DOCUMENT RESUME

ED 266 122 SP 027 191

TITLE	Teacher Evaluation Study. Norming Study, Generic Competence Revision, Feedback Proposal, Cognitive Examinations, and Specialized Domains. Report for 1983-1984.
INSTITUTION	Florida Coalition for the Development of a Performance Measurement System, Tallahassee.
PUB DATE	84
NOTE	283p.; For related documents, see SP 027 189-191.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC12 Plus Postage.
DESCRIPTORS	Classroom Observation Techniques; *Evaluation Criteria; *Performance Factors; Program Development; *Teacher Behavior; *Teacher Evaluation; *Test Norms; Test Validity
IDENTIFIERS	Florida; *Teacher Evaluation Project FL

ABSTRACT

The work of the Florida Teacher Evaluation Project was based upon the assumption that one of the most promising approaches to the improvement of student learning is through the study and modification of teacher behavior. The work has produced products in the areas of domain documents, handbooks, and manuals. This report is an extension of the previous year's work. The establishment of norms and standards of a performance measurement system, and testing the research base and the Florida Performance Measurement System (FPMS) for generalizability were two major objectives of the project. Other aims included continuing development of the research base and observation instruments for evaluating teacher performance in relation to student achievement, and developing and testing a feedback procedure and preliminary examinations for approving observers. Generic competencies were also revised. This final report is divided into five sections: (1) norming study; (2) development of new domains of performance; (3) a proposed feedback system; (4) development of the FPMS cognitive examination; and (5) proposed revision of Florida Generic Competencies. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED266122

TEACHER EVALUATION STUDY

Norming Study, Generic Competence Revision, Feedback Proposal, Cognitive Examinations, and Specialized Domains

SP 025 191

Report for 1983-1984

Florida Coalition for the Development of a Performance Measurement System

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

B.O. Smith

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Conducted Under Special Appropriation 339A
Special Category, Teacher Education Program
from the General Revenue Fund

Table of Contents

		Page
Research/Administrative Personnel and Acknowledgements		iii
Report Summary		vi
Introduction		xi
Section I	Norming Study	1
Section II	Development of New Domains of Performance	31
Section III	A Proposed Feedback System	41
Section IV	Development of the FPMS Cognitive Examination	50
Section V	Proposed Revision of the Florida Generic Competencies	59
Bibliography		67

Appendices:

Section I, A-G	Norming Study: Methods and Tables	1
Section II, A-D	Development of New Domains of Performance	47
Section III, A-B	Proposed Feedback System	120
Section IV, A	Cognitive Examination	132
Section V, A-B	Proposed Revision of the Florida Generic Competencies	135

RESEARCH STAFF

Principal Investigators
Jean M. Borg
Theodore Micceri
Donovan Peterson
B. Othanel Smith

Research Assistant
Melva Underbakke

ADMINISTRATION

Steering Committee
Kathy Bell
Jean M. Borg
Mary Bullerman
Betty V. Fry
John H. Hansen
Louis Morelli
Laurie Murray
Donovan Peterson
B. Othanel Smith
Robert Soar
Benjamin Stephenson
Garfield W. Wilson

ACKNOWLEDGMENTS

Norming Study Observers

Broward County
Beverly Baird
Jacqueline Barber
Frank Berry
Ronald Braynon
Hilton Brown
Sherry Clarke
Mark Gockel
Gail Daly
Jill Dorler
Diane Gordon (coordinator)
Anna Mahoney
Mary Mitchell
Gayle Moller
Jane Muehlberg
Sara Rogers
Ruben Ransaw
Carlos Sasse
Robert Schmelz
Frances Templon
Peggy Truitt
Tony Weisberg
Shirley Woods

Charlotte County
Terry Bohlander
Rene DesJardins
Raymond Jorgenson

Citrus County
Eugene Cram
Joe Crawley
Sandra Cross
Howard Hawk
Jerry Jamison
Sylvia Kohler
Laurie McLennan
Betty Pascall
Phyllis Pullar
Evelyn Richardson
Doris Rooks
Joyce Scott
Carole Slattery
Pat Stanton
Ed Staten
Lane Vick
JoAnn Wagner
Bonnie Wise
Dorothy Zay

Collier County
Charles Ahearn
Michael Green
Henry Jones
Gene Olliff
Santo Pino

Gulf County
David Bidwell

Hernando County
Elaine Beeler (coordinator)
Susan Cannoy
Theresa Gross
Helen Keith
Maureen Richie
Martha Spindler
Michael Tellone
Wendy Tellone
Glenda Vasquez

Hillsborough County
Bernandine Butler
Fran Vanse Evers
Mary Sue Hill
Pearl Hojnacki
Jean Leone
Vincent Nuccio
Charlotte Smedberg

Lee County
Carol Bromby
Steven Cook
Robert Durham
Mary Nell Gunter (coordinator)
Keith Lindahl
Jon Looney
Faye Nalls
Douglas Whittaker

Marion County
Ruth Marcos

Pasco County
Larry Albano
Audrey Joyce
John Joyce (coordinator)
Don McClain
Renee Sedlack
Terry Turner

Sarasota County
Melida Barton
John Bauer (coordinator)
Judy Dunn
Harriet Fisk
Gail Horton
Lucy Laudano
Eloise Malinksy
Denise Maughan
Mike McHugh
Janet Rietz
Pat Sawyer
Jennifer Thatcher
Linda Townsend

Volusia County
Belle Altice
Clare Beaumont
Gwen Carson
Beverly Davies
Mary Lou Deeley
Shirley Edwards
Joseph Faircloth
William Fink
David Fisher
Jerry Gill
Lynn Gold (coordinator)
Al Guenther
Carolyn Harkey
Paige Harper
Nana Hilsenbeck
David Hinson
Cathy Homan
Barbara Krulich
Joe Krulich
Ed Langley
Ronald McPherson
Michael Mogelli
John Nadal
Carl Persis
Cliff Robertson
Margaret Sanchez
Dolores Sapp
Richard Schuler
Virginia Sharpe
Betty South
Don Travis
Mary Wright

Washington County
David Bidwell

COGNITIVE TEST ITEM WRITERS

Charles Ahearn
Jean M. Borg
Mary Bullerman
G. Allen Geyer
Gene Olliff
Donald E. Orlosky
Karen Shapiro
B. Othanel Smith
Nancy Thigpen
Melva Underbakke
Nancy Wilson

FEEDBACK SYSTEM -- FIELD TEST

Kathy Bell
Lore Nielsen

CONSULTANTS

David Berliner
Homer Coker
N. L. Gage
Donald Medley
Joseph Mazur
Steven Phillippy
Robert Soar

Special mention must be made of Susan Allen, Birsen Juston, Pansy DuBose-Pearson and Patricia Zenisek who did typing and office operations that made it possible to carry out this project.

A project of this magnitude required the involvement of a large number of people in a variety of activities. It was impossible to credit all deserving persons but we have listed those names for which we have a record. We apologize for any omissions.

04/10/84
CPT:RS1-3

Summary

This report is divided into five sections:

I. Norming Study

II. Development of New Domains of Performance

III. A Proposed Feedback System

IV. Development of the FPMS Cognitive Examination

V. Proposed Revision of Florida Generic Competencies

I. NORMING STUDY

A norming study of the FPMS summative observation instrument was conducted during late fall and winter of 1983/1984. Normative data were obtained using cluster sampling in which total schools were used as sampling units. One thousand two hundred and twenty-three observations were conducted at all grade levels and in all subjects by trained observers in 45 schools and 13 counties. The sample was selected to be representative of Florida's schools using criteria such as: urban/rural, student SES, grade levels, subjects, and teacher experience.

Scaled scores based on semi-interquartile ranges derived from the normative data were tested for reliability using the data obtained during the 1982-83 reliability study and were found to produce estimates as high or higher than those of the normalized scores originally used in the study (effective total score reliability estimate r=.91; ineffective total score reliability estimate r=.84).

Norming study results provide support for the following conclusions:

1. The instrument appears to clearly separate effective teachers, as

defined by the domain documents of the FPMS, from ineffective teachers within all groups.

2. Except for differences in scores between elementary school teachers and middle/secondary teachers, there appear to be no major differences between group mean scores on the instrument. All ethnic groups, experience levels, education levels, types of student groups, types of classroom conditions, etc., had similar mean scores.

These results indicate that the FPMS summative instrument satisfies the following key objectives of any teacher evaluation instrument: content validity, concurrent validity, reliability, feasibility, scalibility, flexibility, discrimination, feedback, and relevance. The FPMS summative instrument may be used for evaluation, feedback, and remediation.

In using data from FPMS observations for purposes of evaluation, we recommend that the semi-interquartile range be used for scoring and that superior performance be defined as those persons scoring above the median on both the effective and ineffective indicators. This allows ranking of merit applicants relative to the norms.

II. DEVELOPMENT OF NEW DOMAINS OF PERFORMANCE

During 1983-84, three FPMS domains of research findings were compiled to extend the original set. The focus of each new domain was somewhat different, i.e., Domain 6.0 on Testing included what the effective teacher does in the classroom prior to testing as well as during and after testing; Domain 7.0 is a specialized domain because, unlike all the others, it deals with teacher behavior in a one-on-one setting; and Domain 8.0 is unique in

that the focus is on how to help the teacher.

III. A PROPOSED FEEDBACK SYSTEM

Each year the public schools of Florida employ several thousand beginning teachers. In 1983-84, approximately 4500 beginning teachers were employed. Some 2000 or about 45 percent were graduates of Florida institutions of higher education. The Florida statutes require that a support team be assigned to each beginner to help him/her adjust to the job and improve teaching performance. The feedback procedure herein proposed is designed to collect data on the performance of beginning teachers and their opinions of how well their preservice preparation matches the requirements of their job. The usefulness of this information is found in the need to:

--develop and monitor support team programs,

--set minimum performance levels for certification of beginning teachers,

--encourage program improvement in teacher education colleges and academic departments,

--stimulate inservice program development for beginning teachers,

--provide information to school districts about the quality of out-of-state graduates and the institutions from which they came,

--identify problems of the Beginning Teacher Program, and to design studies relevant to them.

A feedback system is proposed with the objectives above in mind, a system that places the responsibility for collecting, processing, and distributing information about the Beginning Teacher Program in the State Department of Education.

IV. DEVELOPMENT OF THE FPMS COGNITIVE EXAMINATION

During 1983-84 the corps of FPMS trainers have trained about 2400 people to be observers using FPMS instruments and materials. To this end, the Florida Coalition has pursued a thorough training program. It is equally important to measure the effect of that training on the quality of observations. This year we have developed and used a comprehensive testing and evaluation procedure. The development of 500 test items, from which various test forms can be generated, will facilitate the assessment of observer competence in all applications of the FPMS.

V. PROPOSED REVISION OF FLORIDA GENERIC COMPETENCIES

The original competencies were incorporated into State Board of Education rules for teacher education and certification in 1978. Ten new competencies have been gleaned from recent research. These are proposed as additions to the 24 original competencies.

These new competencies, like those in the original list, are not strictly speaking competencies but rather clusters of teaching skills. In developing the original list, it was early recognized that the competencies were too general for use in either test construction or classroom observation. Consequently, they were analyzed into subskills. However, their high inference level tended to preclude their use in observing teacher performance. The proposed revision replaces the original subskills with low inference skills that specify teacher behavior. This makes valid and reliable classroom observations feasible. In addition, the competencies have been classified into three groups: those most appropriately evaluated 1) by cognitive tests, 2) by interview procedures or analysis of products,

and 3) by observation of teacher behavior in the classroom.

04/10/84
CPT:S1-4

Introduction

The present research and development on teacher evaluation, provided for in Section 231.87 of the Florida Statutes, began in the academic year 1982-1983.

This work was carried out under the auspices of the Florida Coalition for the Development of a Performance Measurement System, a coalition organized and directed through the Office of Teacher Education, Certification and Inservice Staff Development, State Department of Education.

From the beginning, the Teacher Evaluation Project assumed that one of the most promising approaches to the improvement of student learning is through the study and modification of teacher behavior. It is well documented that teacher behavior is significant in the learning process (Soar, Evertson, McDonald, Stallings, etc.). The work of the Teacher Evaluation Project has been consistent with this assumption.

The products of this work (e.g. domain documents, handbooks and manuals) have been made available through the Panhandle Area Education Cooperative print shop. To date, over 17,250 copies of these products have been purchased by school districts and universities within the State and over 800 copies have been purchased by out-of-state institutions.

The present report, 1983-84, is an extension of the previous year's work. In accord with the plan submitted to the Commissioner of Education pursuant to Specific Appropriation 339A (Special Categories, Teacher Evaluation Program, from the General Revenue Fund), the work during 1983-84 was conducted in the following areas:

1. Establish norms and standards of a performance measurement system,

and test the research base and the FPMS Summative observation instrument for generalizability.

2. Continue development of the research base and observation instruments for evaluating teacher performance in relation to student achievement.
3. Develop and test a feedback procedure.
4. Develop preliminary examinations for approving observers.
5. Revise generic competencies.

The work completed in each of these general areas is described in the various sections of this report.

Section I

NORMING STUDY

The Florida Performance Measurement System (FPMS) was developed from process product and experimental research on teacher effectiveness. The literature reviewed by the research team of the Coalition produced a total of 31 concepts and 124 indicators of teacher effectiveness. This information was compiled into six domain documents and bound under the title Domains: Knowledge Base of the Florida Performance Measurement System.

These researches were conducted at specific grade levels and in specific subjects. Many were conducted in elementary schools, while others were conducted in secondary schools. For example, The Beginning Teacher Study (McDonald, 1978) was conducted in reading and mathematics in the second and fifth grades. Some claim that the concepts and indicators of teacher effectiveness likely apply only to teaching contexts in which they were studied. Our research problem was to determine (1) if these behaviors exist in all teaching contexts, and, if so, (2) whether any differences in these behaviors exist in varying contexts. We sought to identify "normal" teacher behavior.

Norming is fundamental to converting a measurement instrument, such as the FPMS, to an evaluation system. More specifically, our purposes in norming were:

1. to provide a basis for setting preliminary norms based upon observed behaviors of teachers in practice;
2. to identify factors--grade level, subject, classroom conditions, student variables, teacher variables, and instruction format that cause different frequencies of generic behaviors in teacher presen-

tations;

3. to determine whether variables that significantly impact student achievement also alter teacher behavior in varying contexts;
4. to identify clusters of teacher behavior common to specific contexts of teaching;
5. to develop the basis and rationale for using summative instrument scores for purposes of evaluation.

Preparations for the Norming Study

The use of any instrument and/or procedure for purposes of evaluating personnel is dependent upon evidence that the instrument and procedure accurately measure the phenomena being evaluated. Applied to teacher evaluation, this means that the instrument must be valid and reliable, the observers must be reliable, and the contexts in which the instrument is to be applied must be tested to determine if different norm groups exist.

The work done last year established the content validity of the FPMS, i.e., the concepts and indicators were drawn from research on teacher effectiveness and knowledgeable people agreed that they were defined, grouped, and represented accurately. Reliability coefficients were established for interrater agreement--the degree to which two or more observers working independently agree upon the recording of an indicator of teacher behavior ($r=.85$); for stability--the degree to which observers consistently record an indicator of teacher behavior when displayed on two or more occasions ($r=.86$); and for discriminant reliability--the degree to which the instrument consistently ranks teachers on a scale of effectiveness ($r=.79$). These

coefficients were obtained using percentile based normal standardized scores. Such tests were prerequisite to the norming study, for irrelevant items and observer errors can make evaluations of teacher performance meaningless.

Research on instrument validity and reliability was presented in the 1982-83 report. The current year's work began by identifying the study sample and testing the observation accuracy of the persons who volunteered to make the hundreds of observations required for the norming study.

Sample Selection

Factors contributing to sample selection of schools for the norming study included size (number of students), grade level, socio-economic status of the students, subject area, and whether the schools were located in urban or rural areas. Participation in the norming study was voluntary. The Coalition served as the medium for explaining the purposes of the norming study and soliciting participation of observers. Forty-five schools in the following 13 districts volunteered to participate:

Broward	Gulf	Marion	Washington
Charlotte	Hernando	Pasco	
Citrus	Hillsborough	Sarasota	
Collier	Lee	Volusia	

These districts vary in size, urban/rural location, and socio-economic status. They were found to adequately represent Florida's school population. Of the schools observed, size ranged from 3,000 students in Leto High School, Hillsborough County, to 370 students in Floral City Elementary

School in Citrus County. Socio-economic status of the students ranged from 2% to 84% low SES as determined by data from the State's Free-Lunch Program.

Observations of 468 elementary teachers (grades K-5) were conducted in 17 schools, 226 middle school teachers (grades 6-8 or 9) in 9 schools, 520 high school teachers (grades 9 or 10-12) in 14 schools, and 11 teachers in a Center for Adjustment.

Cluster sampling--observation of all teachers in each school--was used to assure that all subject areas would be represented. A total of 1223 observations were conducted by 117 trained and tested observers. (See acknowledgments for a list of persons who conducted the observations.)

Many of the observers had been trained previously and had conducted observations using the FPMS summative instrument in the Beginning Teacher Program. All observers used in the norming study were given either a training review or a standard three-day training program designed by the Coalition and conducted by approved trainers. All observers used in the study passed an observation reliability check. This was done using prerecorded videotapes of regular classroom instruction. Reliability observations were scored by the research team.

Observations for the norming study were scheduled and conducted during October, November, and December of 1983. All data were collected using standard FPMS instruments and forms. These included the FPMS Summative Instrument, and Frame Factor and Instruction Format forms that were used to collect information on independent variables such as grade level, subject, class size, SES, type of lesson, etc. (See Appendix Section I, A for copies of these forms and instruments.) Frame Factor and Instruction Format forms

were completed by the observers and teachers with assistance, when necessary, from the principal or other staff personnel. Completed instruments and forms were then collected by district coordinators and transmitted to the research team for data entry and analysis. Data entry and initial data analyses were conducted during December, 1983 and January, 1984.

Demographic Description

The information reported in this section is a quantitative description of the independent variables (frame factors) included in this study. The representativeness of the norming study sample as it relates to all of public education in Florida may be judged from these data.

The following pages describe the characteristics of the schools, teachers, students, and lessons observed. Appendix Section I, B contains detailed tabular information for all the variables (frame factors) discussed below for which tables are not reported in this narrative.

Characteristics of Schools

All subjects taught in elementary and secondary education in the sampled schools were classified initially into nine major categories. (See Table 1.)

Table 1

Major Subject Categories

Category Label	Absolute Frequency	Absolute Percent	Relative Percent
Language Arts	412	33.7	34.9
Math	174	14.2	14.7
Science	122	10.0	10.3
Social Sciences	123	10.1	10.4
Arts - Fine & Practical	240	19.6	20.3
Special Ed.	75	6.1	6.4
Foreign Language	31	2.5	2.6
Guidance	2	0.2	0.2
Other	2	0.2	0.2
Missing Information	42	3.4	Missing
Total	1223	100.0	100.0

The most commonly occurring subject was Language Arts, comprising 34.9% of all observations. Several subject areas were represented by fewer than 50 observations, including all specific subcategories of the skill-based Arts (Fine and Practical) detailed in Table 2.

Table 2

Subject Subcategories--Arts (Fine and Practical)

Category Label	Absolute Frequency	Absolute Percent	Relative Percent
Art, Drama, Speech	19	1.6	9.0
PE	43	3.5	20.5
Music	28	2.3	13.3
Agriculture	7	0.6	3.3
Business-Marketing	23	1.9	11.0
Cleric-Sect-Data	34	2.8	16.2
Industrial-Dct-Shop	20	1.6	9.5
Crafts-Home Ec	31	2.5	14.8
All Other Subjects (Academic)	1018	83.2	Missing
Total	1223	100.0	100.0

Table 2 shows, for example, that a total of 28 observations were made of music teachers (13.3% of the subgroup and 2.3% of all observations in the total study).

All grade levels were represented in the study (Table 3). However, secondary schools were oversampled to obtain more information regarding the specialized subjects in secondary schools. As an example, 188 observations were conducted in the 11th grade (some of which crossed grades 11-12), while only 51 were conducted in the 6th grade.

Table 3

Grade Levels

Grade Level	Absolute Frequency	Absolute Percent	Relative Percent
0 (Kindergarten)	60	4.9	5.0
1	78	6.4	6.5
2	74	6.1	6.1
3	92	7.5	7.6
4	91	7.4	7.5
5	73	6.0	6.1
6	51	4.2	4.2
7	71	5.8	5.9
8	82	6.7	6.8
9	115	9.4	9.5
10	140	11.4	11.6
11	188	15.4	15.6
12	91	7.4	7.5
-9	17	1.4	Missing
Total	1223	100.0	100.0

Nine different types of classroom settings were observed, of which the self-contained type is by far the most common (82%). Other types were pods, resource rooms, physical education fields, labs, media rooms, etc..

Teacher perceptions of classroom conditions were also obtained. Each teacher rated his/her facility as being poor, adequate or excellent with regard to such factors as crowding, lighting, noise, etc.. Ninety-two percent of the teachers rated their classrooms as adequate or better.

Characteristics of Students

Classes ranged in size from a single student to over 100 students. Median class size was 24.46 (mean=23.68). The average class, consisting of 23.68 students, contained 12.2 male students and 11.4 female students.

The majority of teachers (78%) reported no gifted students, and 54% of the teachers reported no students with learning impairments. Thus, the average class contained fewer than a single (.79) gifted student, and less than two (1.779) individuals having a learning impairment.

One of the primary predictors of student achievement in the research literature is socio-economic status (SES). This is a difficult variable to measure in practice, as no clear definitions of the variable's parameters exist. In this study, two measures were used to tap this factor: teacher perceptions of student SES, and data from the state free lunch program (1982-1983). These two measures of SES correlated positively (r=.61). Both indicated that approximately 30% of the students could be defined as "Low SES." This percentage is slightly below the statewide average (33.5% on free lunch) largely as a result of the oversampling of high schools which tend to have fewer students on free lunch than either elementary or middle schools.[1]

---

[1] One high school contributed 107 observations to the sample, while having only 2% of its students on free lunch.

Information was also obtained on the English language ability of students. Student speech was classified as: (1) standard English, (2) a foreign language only, (3) English as a second language, and (4) a non-standard dialect. Very few classes contained students who speak non-standard English. An average of fewer than one student per class speaks a foreign language, English as a second language, or a non-standard dialect of American English.

Teacher Characteristics

As with the other categories in the demographic description of the norming study sample, information was collected on teachers to determine which, if any, characteristics may be significant in establishing normed groups for the FPMS summative instrument. Teacher attributes included highest degree (Table 4), sex (Table 5), ethnicity (Table 6), and number of years of teaching experience (Table 7).

Table 4

Highest Degree

Category Label	Absolute Frequency	Absolute Percent	Relative Percent
Non-Degree	10	0.8	0.8
B.A., B.S., B.F.A.	675	55.2	56.7
M.A., M.S., M.Ed.	467	38.2	39.2
Ed.S., Ed.D., Ph.D.	38	3.1	3.2
Missing	33	2.7	Missing
Total	1223	150.0	100.0

Only ten teachers lacked college degrees (trade and industrial teachers). Most teachers held Baccalaureate degrees (55%) and a substantial number held Master degrees (38%).

Table 5

Sex of Teachers

Category Label	Absolute Frequency	Absolute Percent	Relative Percent
Male	351	28.7	29.1
Female	857	70.1	70.9
Missing	15	1.2	Missing
Total	1223	100.0	100.0

Approximately 71% of the teachers in the sample were women while only 29% were men. This is attributable partly to the large number of female teachers in elementary schools.

Teacher ethnicity in the sample was 86.9% Caucasian, 10.5% black-non-Hispanic and 2.2% Hispanic, as noted in Table 6.

Table 6

Ethnicity of Teachers

Category Label	Absolute Frequency	Absolute Percent	Relative Percent
White non-Hispanic	1038	84.9	86.9
Black non-Hispanic	125	10.2	10.5
Hispanic	26	2.1	2.2
Other	6	0.5	0.5
Missing	28	2.3	Missing
Total	1223	100.0	100.0

Years of teaching experience ranged from less than one (30 cases) to thirty-seven (3 cases). The mean ($\bar{x}$) level of experience was approximately twelve years.

The typical teacher in this study was a white female with a Baccalaureate degree and twelve years of experience.

Instruction Format

Teachers have little control over the demographic variables described above. We call these Frame Factors, or the context within which the teacher functions. Instruction Format variables differ from the Frame Factors listed above in that teachers have some choice in their selection. These include grouping of students, instructional method, lesson type, and length of lesson.

Three methods of class grouping for instruction were total group, subgroups/independent instruction, and a combination of the two. Table 7 shows that most of the observations were of total group instruction (745 of 1223) while subgroup/independent instruction and combinations of total and subgroup/independent instruction were observed equally (196 observations in each).

Table 7

Grouping

Category Label	Absolute Frequency	Absolute Percent	Relative Percent
Total Group	745	60.9	65.5
Subgroup or Ind.	196	16.0	17.2
Both	196	16.0	17.2
Missing	86	7.0	Missing
Total	1137	100.0	100.00

Teaching method in this study was recorded as lecture, discussion/ /interaction/ recitation, independent work, or any combination thereof. Over 80% of all classes observed involved some form of discussion/interaction/ recitation. (See Table 8.)

Table 8

Teaching Method

Category Label	Absolute Frequency	Absolute Percent	Relative Percent
Lecture (1)	69	5.6	6.2
Disc/Inter/Recit (2)	318	26.0	28.6
Independent (3)	102	8.3	9.2
1 + 2	350	28.6	31.4
1 + 3	23	1.9	2.1
2 + 3	153	12.5	13.7
All Types	98	8.0	8.8
Missing	110	9.0	Missing
Total	1223	100.0	100.0

We initially hypothesized that the summative instrument would code differently depending upon the type of lesson being taught. Data were, therefore, gathered on whether observations were conducted during an introductory or developmental lesson, a review or practice lesson, a combination of the two, or during a test. (See Table 9.)

Table 9

Lesson Type

Category Label	Absolute Frequency	Absolute Percent	Relative Percent
Intro or Developmental	325	26.6	28.8
Review or Practice	377	30.8	33.4
Combination	413	33.8	36.6
Test	13	1.1	1.2
Missing	95	7.8	Missing
Total	1223	100.0	100.0

The most frequent lesson observed included both introductory/developmental and review or practice, i.e. diversified lessons (413 observations). Introductory/developmental lessons constituted 26.6% of the observations and review/practice constituted 30.8%. Observations during administrations of tests were conducted on thirteen occasions.

Observers were instructed to code an entire lesson from start to finish. Table 10 shows variations in length of the classes observed in the norming study. These include 15-29 minute classes (118 observations), 30-39 minute classes (548 observations), 40-49 minute classes (228 observations) and 50-60 minute classes (329 observations).

Table 10

Length of Observations

No. of Minutes		Absolute Frequency	Absolute Percent	Relative Percent
15-29		118	9.6	9.6
30-39		548	44.8	44.8
40-49		228	18.6	18.6
50-60		329	26.9	26.9
	Total	1223	100.0	100.0

The longest classes observed were sixty minutes. Since the FPMS is a frequency-based system, variations in observation time are corrected as follows: (item score/length of observation) X median length of observation).

All of the afore-mentioned variables were used in analyses of scores on the summative instrument item frequency.

Scaling

Two methods of scaling the instrument were tested; 1) standardizing observations, and 2) percentile-based scores. In standardizing, absolute raw scores, by item, were transformed to become "normal standardized scores." This procedure reduces the detrimental effects of long-tailed distributions on the resulting analysis--a phenomenon common to observation distributions. (See Appendix Section I, C for detail.) Normal scores were

then standardized to a mean of five and a standard deviation of one (20 effective items x5=100).

This procedure posed two serious problems that we were not able to resolve. First, if applied to merit evaluations, it would require each observer to code a minimum of ten teachers preferably in different subject areas and at different levels. Current rules for merit require one or more of the observers to be a subject specialist, and one observer must be from another district. It is, therefore, impractical to expect all observers to code ten teachers. Second, this procedure rank orders subgroups of teachers (those observed by one or more persons) independently of all other groups of observed teachers. Within any given group, some will score high while others will score low. This leaves open a remote possibility that a teacher might score considerably lower against one group than s/he would against another group. This could eliminate a deserving teacher from merit. For these and other reasons, a second scaling technique was tested--quartile-based scores.

The Florida State Legislature requires (Florida Statutes 231.0881) that merit scores be related directly to a normed population. The major advantage of quartile-based scaling is the direct relationship it allows between individual scores on the FPMS summative instrument and the normed scale. This provides a basis for judging the quality of teaching, as defined by the summative instrument, over a number of years. The base year, 1983/1984, serves as the standard against which subsequent years may be compared.

The normative study indicates that several items on this instrument occur infrequently (50% of the positive items have medians below 1.50). In

addition, some observers code more frequently than others--a matter of pacing. Given these factors, frequency-based scaling would tend to over-weight the more frequently occuring items. For these reasons, a three point scale for each item based upon the semi-interquartile range was tested as a basis for scaling. An individual in the top 25% of frequency for an item receives a score of 3 points, in the 25% to 75% range a score of 2, and in the bottom 25% a score of 1 point. Individual scale points are based on fractions (the percentage of a point that 25% of the teachers comprise), as all item scores are fractions following correction for observation length. In computing reliability, and in future evaluations, an individual item score will be the mean produced by three observers, again, a fraction. These fractions may be considered "True Scores." (For greater detail on scale development see Appendix Section I, D.)

Using this method, teachers scoring high tend to be those exhibiting multiple behaviors in many categories, while those scoring low exhibit few behaviors in few categories. This scaling was then applied to the reliability data (1982/1983 report) with positive results. It appeared to increase the variance among teachers (ranking them in almost exactly the same order as the original standard normalized method), without simultaneously increasing the variance either among teams of observers or lessons.

Again, this scaling seems to satisfy the criteria specified by the legislature. It is directly related to the normative data, it appears valid for the instrument's purposes (many behaviors, many categories - high scores; few behaviors, few categories, - low scores), and it shows a reliability level equal to or greater than reliability obtained from the

standard normalized scores. (Sees Tables 11 and 12.)

Table 11

Reliability Estimates for Five Separate Scales
Using Percentile Based Scores

Effective Indicators

Type of Reliability	Total Scale 20 Items	Domain 2 4 Items	Domain 3 9 Items	Domain 4 4 Items	Domain 5 3 Items
Discriminant	.91	.77	.90	.94	.52
Stability	.92	.77	.90	.95	.56
Interrater	.98	.94	.98	.98	.83

Table 12

Reliability Estimates for Five Negative Scales
Ineffective Indicators

Type of Reliability	Total Scale 20 Items	Domain 2 4 Items	Domain 3 9 Items	Domain 4 4 Items	Domain 5 3 Items
Discriminant	.84	.74	.86	.17	.91
Stability	.85	.75	.87	.18	.91
Interrater	.98	.98	.98	.97	.98

Given these advantages of the percentile-based method described above, and the relatively high estimates of reliability, the norming data were subsequently analyzed using this method.

Data Analysis

As stated earlier, observations were obtained for 1223 teachers, observed by 117 trained and tested observers in 41 schools (17 elementary, 9

middle and 14 high schools) in 13 counties around the state of Florida. Quartile-based scores were used in the analyses reported herein.

Two procedures (regression analysis and factor analysis) were used to determine the degree to which the FPMS summative instrument is generic and generalizable to K-12 teaching contexts, and to determine the number and nature of possible norm groups required when using this instrument.

Regressions were run separately from all continuous independent variables (all student variables, most school and classroom variables, and teacher experience) to each item on the summative instrument, to a total score of each domain (positive and negative), and to each positive and negative total score. These analyses were conducted first upon one randomly selected sub-sample (n=670) from the total group. Significant and meaningful results were then cross-validated upon the second sub-sample (n=553). The purpose of these analyses was to determine which if any of these variables may cause consistent variations in teacher behavior on the summative instrument and, therefore, call to question the need for separate norm groups or limits of generalizability.

For all categorical independent variables (grade level, lesson formats, subject areas, classroom type, and educational level) ANOVA's were run first on sample 2 (n=553) and then tested in a factorial design on sample 1 (n=670) against the hypothesized most important factor (lesson format: instructional method) to determine if main effects existed separately for the independent variables.

In these analyses, three considerations were of importance:

1. Since several ANOVA's were being run, the significance level

acceptable for continued considerations was set at .01.

2. In addition, due to the large n's (500 or more in some analyses) the meaningfulness of differences determined to be significant is important. ANOVA's run on 500 or more cases tend to have significant results for means that are only minimumly different. A criteria of at least 1/3 standard deviation, or 2.15 points difference between groups for the effective total score was set as a minimum for further consideration. This assures that a mean difference of only a single point on a total score will not put individuals in different norm categories.
3. At the other end of the continuum, some of the subgroups in the analysis had very small n's. Unusually high or low frequency observations tend to skew distributions when this is the case. Therefore, at least 75 observations had to be obtained for any given subgroup in order to accept results as valid for norming purposes.

Norming Study Results

This section describes the relationship of independent variables (e.g. grade level, subject areas, sex, ethnicity) to the dependent variables (effective and ineffective scores on the FPMS summative instrument). This relationship is determined by similarities or differences of dependent variable distributions across subgroups of independent variables. Differences in mean scores or distribution shapes are the basis upon which normed groups are created.

Applying quartile-based scores (described in data analysis) to the norming observation scores on the effective side of the summative instrument allows a range from 20 to 59. Scores on the norming observations actually ranged from 20 to 57. Scores on the ineffective side of the instrument can range from 19 to 38. Actual range in the norming study spanned this entire range.

A criteria of .33 standard deviation from the grand mean score (S_X = 6.86 X .33 = 2.15) was set as the acceptable level of variation within norm groups. A minimum sample size of approximately 75 had to be obtained for consideration as a separate norm group. These criteria were set to avoid false effects of outlying observations on any given distribution caused by a small n in the subgroup. The findings of this study were, therefore, analyzed by first considering sample size (75 or more observations) and then differences in subgroup mean scores to see if they exceeded .33 standard deviation (2.15 points). Then, if other factors were similar, either separate norm groups were created or subgroup scores were adjusted using a standard factor. (See Appendix Section I, E.)

The first finding of the study was that teachers on an average were coded in far fewer ineffective categories (1.5) than they were coded in effective categories (14.3). However, the range of ineffective scores allows for high level of discrimination. Some teachers, although very few, demonstrate high frequencies of ineffective behavior while most demonstrate little or no ineffective behavior. Using the criteria set above, there is no evidence that separate norm groups should be created for the ineffective side of the instrument, nor are standard factors required to adjust subgroup

scores. All teachers appear to be part of the same norm group on the ineffective side of the summative instrument. (See Appendix Section I, E for scores within subgroups on the ineffective side of the instrument.)

Two factors emerged as being of primary import to norming on the effective side of the summative instrument--grade level and instructional method. Analysis of grade level showed significant differences between elementary and secondary teachers as they were coded on the effective side of the instrument.

The second factor of importance in norming proved to be instructional method. Observations coded on the effective side of the summative instrument varied significantly depending upon teaching method, i.e. lecture, interaction, or independent seat or lab work. Rather than creating separate norm groups, score adjustments for the two lower scoring methods proved to control adequately for this factor. (See Appendix Section I, F.) Thus, only two norm groups result (elementary grades K-5 and post elementary grades 6-12).

Analyses were conducted on all independent variables. (See Appendix Section I, A for a list of independent variables.) Results of these analyses are reported below and in Appendix Section I, G.

Independent Variable Analysis

Grade level, as it relates to the total score, was the most important independent variable in creating norm groups on the effective side of the instrument. (See Table 13.)

Table 13

Group Scores By Combined Grade Level

Level	Mean	Std. Dev.	N
Elementary (0-5)	38.55	6.07	468
Post Elementary (6-12)	35.50	6.70	721
Total	36.70	6.63	1189

An analysis of Table 13 shows that both criteria for creating norm groups are met--there is a 3.63 point difference in mean scores and both groups far exceed the minimum n of 75. Data were combined into three grade levels in the norming study (elementary, middle and secondary); however, the scores show that middle and secondary teachers appear to be part of the same group while elementary teachers form a separate group. Differences in mean scores (elementary scoring higher than secondary) appear to be attributable to two main factors. First, elementary students require more specific instruction in the form of teacher direction than secondary students and, second, elementary teachers tend to use student/teacher interaction almost exclusively as their instructional method. (Ninety-two percent of all elementary teachers observed in the norming study used interaction as the main method of instruction during observation, while almost 25 percent of all secondary teachers used pure lecture or independent work as their instructional method.)

Small n's in subgroups (less than 75) were obtained in several cases, even though total sample size exceeded 1200 observations. General groupings, therefore, were used in this study to allow statistical analysis. Table 14 shows total, adjusted, positive (effective) scores for major subject areas.

Table 14

Total Positive Scores Across Subjects by Norm Groups

Subject Area	Elementary Mean	n	Post Elementary Mean	n
Language Arts	38.52	240	35.83	170
Math	37.85	61	35.05	110
Science	36.53	19	35.36	100
Social Sciences	39.18	26	35.01	93
Practical Arts	37.91	50	35.23	181
Special Education	39.85	37	36.03	32

Attempting to examine data by dividing general subject areas into more specific subjects usually reduced subgroup sizes far below criterion level. Table 15 is an example. The Arts, Fine and Practical, proved a most difficult area to define, since they include a range of subjects from drama to home economics. All subjects in this major group are skill-based rather than theoretical in nature.

Table 15

Total Positive Scores Across Arts Subjects by Norm Groups

Subject Area	Elementary Mean	n	Post Elementary Mean	n
Art, Drama, Speech	38.24	7	33.13	11
Physical Education	36.48	16	33.03	27
Music	35.18	11	36.36	17
Agriculture	-----	--	37.10	7
Business/Marketing	40.50	2	36.34	19
Clerical/Sect/Data	37.51	6	34.98	26
Industrial/DCT/Shop	-----	--	35.78	18
Home Ec./Crafts	-----	--	32.93	29
Other	44.00	3	37.50	3

Note that the most observations made of one subject was 27 in physical education. Two or three unusually high or low individual scores can inordinately influence a group mean score when the n is this low. Practicality in this research study precluded conducting 75 or more observations in each subgroup. Furthermore, the similarity of the mean scores shown in Table 15 is a strong indication of the generic nature of this instrument.

Variations in types of classrooms were also examined in the norming study. (See Table 16.) The most common facility is the self-contained classroom. Observations were conducted in 339 self-contained elementary classrooms with an adjusted mean score of 39.05, and 625 observations were conducted in self-contained secondary classrooms with a mean score of 35.66.

Table 16

Total Positive Scores Across Classroom Types by Norm Groups

	Elementary		Post Elementary	
Classroom Type	Mean	n	Mean	n
Self-Contained	39.05	339	35.66	625
Pod or Open Class	36.68	70	35.41	14
PE Court or Field	35.12	14	35.14	20
Lab	37.50	2	36.31	16
Resource Room	40.34	15	35.20	7
Departmental	40.97	9	32.59	7
Media/Clinic	38.40	5	33.50	2
Other	35.33	3	30.57	3

Note that the adjusted scores are relatively close within norm groups, except for cases having small n's. The only classroom type, other than self-contained in this Table, that is close to meeting the minimum n criterion is the pod or open classroom, and the difference from the grand mean for

that classroom type, 1.72 points, is well within the criterion of 2.15 points.

Teacher Characteristics

Analyses revealed no significant relationships between scores on the FPMS summative instrument and the following teacher variables: education level, experience, sex, or ethnicity.

The relationship between a teacher's education level and effective scores on the instrument were non-significant as shown in Table 17.

Table 17

Total Positive Scores Across Teacher's Degree by Norm Group

Highest Degree Attained	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
Non-Degree	----	--	35.43	8
B.A., B.S., BFA.	38.49	290	35.63	368
M.A., M.S., M.Ed.	38.73	165	35.39	289
Ed.S., Ed.D., Ph.D.	37.63	8	34.85	30

Years of teaching experience show no relationship to effective scores either; however, a trend, although statistically non-significant and of small magnitude, is evident and interesting because of the steady drop in mean scores as experience increases.

Table 18

Scores Across Teacher Experience by Norm Groups

Years Experience	Elementary Mean	n	Post Elementary Mean	n
One year or less	40.02	59	36.36	33
2 to 4 years	39.21	66	36.41	83
5 to 9 years	38.80	104	35.59	141
10 to 14 years	38.67	125	35.68	199
15 or more years	37.83	141	34.98	241

Examination of the general distribution from which these data were taken shows a wide range of scores in the early years of teaching that steadily narrows with experience. Both low and high scorers tend to disappear over time, with a greater attrition among higher scoring teachers. Part of this trend may be attributed to unsuccessful teachers dropping out or being released from their positions and successful teachers being moved to administrative positions or leaving the profession for more lucrative positions.

Our analyses included consideration of mean differences among the three major ethnic groups in Florida--white, black and Hispanic. Table 19 shows

Table 19

Total Positive Scores Across Ethnic Groups by Norm Groups

Ethnic Group	Elementary Mean	n	Post Elementary Mean	n
White non-Hispanic	38.68	376	35.72	635
Black non-Hispanic	38.28	65	33.04	58
Hispanic	38.00	14	37.36	11
Other	35.00	3	35.00	3

no difference in mean scores among these three ethnic groups at the elementary level. At the secondary level, Hispanics obtained a mean score 2.66 points above the normed mean score of 34.70 and blacks scored 2.49 points below the normed mean score of 34.70. Concern about possible racial discrimination caused us to examine the low scores for black secondary teachers in more detail.

As a result of further analyses, three factors appeared to have some effect on the scores of black teachers in the norming study--sex, grade level, and SES of classrooms. Table 20 shows that the performance of black males was lower than that of black females for both elementary and post elementary levels. This is also true among whites. (See Table 21.)

Table 20

Positive Total Scores for Blacks Across Sex and Norm Group

Subject Area	Elementary Mean	n	Post Elementary Mean	n
Males	36.90	10	31.36	22
Females	38.56	55	34.32	36

Black females score at or about the mean for all teachers in both elementary and post-elementary norm groups. In the elementary schools, black males perform within a reasonable range of the norm mean. However, in the high schools, black males perform below the mean by about three points, and this does not appear to be a result of any combination of factors in the data base, except perhaps the very small number of male black teachers in the sample.

In addition, black teachers in secondary academic classrooms appear to be the low scorers, while black teachers in other areas, such as vocational, music, physical education appear to score on a level with other teachers in post elementary classrooms. Thus, only one small subgroup of black teachers investigated, post elementary males teaching in academic classrooms, appears to account for this seeming difference in scores. We therefore concluded that the differences in mean scores described above are probably due to the sample group. We further conclude that the recommended method of scoring the FPMS summative instrument does not negatively discriminate against any ethnic group. Table 21 shows that although female teachers score slightly higher than male teachers at both levels, the differences are minimal.

Table 21

Total Positive Scores Across Sex by Norm Groups

Sex	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
Male	37.76	53	35.27	285
Female	38.67	410	35.65	429

Student SES

A major finding in the research literature on teacher effectiveness addresses differences in effectiveness of similar teacher behaviors (high and low order questioning is an example), depending upon the socio-economic status of students. The last table addressed in this narrative concerns the relationship between student SES and mean scores on the FPMS summative instrument. As can be seen on Table 22, there is no consistent relationship

between the SES of students and teacher's total scores.

Table 22

Total Positive Scores by SES Characteristics by Norm Groups

Percentage of Low SES Students	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
20% or less low SES	38.11	180	35.42	368
20% to 39% low SES	38.77	128	35.82	87
40% to 59% low SES	40.23	27	33.82	37
60% to 79% low SES	41.70	11	37.97	18
80% or more low SES	38.97	96	34.37	105

This should not be interpreted as being contradictory to research findings on teacher effectiveness. We are reporting total scores. Teachers could, for example, have emphasized the use of high order questions with high SES students and low order questions with low SES students in the study sample. Total scores do not bias a teacher's score because of such differences.

Tables comparing other independent variables (frame factors) with observation mean scores may be found in Appendix Section I, G. Analysis of all of these tables has led us to conclude that the FPMS summative instrument generalizes across all frame factors and that by correcting for instructional method, only two norm groups appear (elementary and post elementary) using the FPMS summative instrument on K-12 teachers in Florida.

From Measurement to Evaluation

Following development and testing, the FPMS summative instrument was determined to be valid and reliable for measuring teacher classroom perfor-

mance. Completion of the norming study has provided the basis upon which data, collected using this measurement instrument, can be converted to scores for evaluating teachers. This claim is based upon results of a norming study that supports the following conclusions:

1. The majority of teachers regularly use several of the indicators on the FPMS summative instrument in their daily classes (75% use 12 or more in a lesson).
2. All groups and subgroups of teachers show considerable variation in the use of both positive and negative indicators. This is true with regard to both the number of indicators and the frequency of specific indicators.
3. The generic nature of these behaviors across all forms of teaching appears strongly supported.
4. It is possible to compare the frequency and distribution of behaviors exhibited by an individual teacher to those found among the population of Florida teachers.
5. Based upon the foregoing points, a teacher's behavior can be evaluated against that of the population of Florida teachers.

Considering the preceding propositions, sufficient basis exists for the evaluation of individual teacher performance within the limits of measurement error for this instrument.

It is recommended that the semi-interquartile range as described in this report be used for scoring. Individuals are thereby scored as being in the lowest 25%, in the mid 50%, or in the highest 25% of the normed population with regard to each item on the instrument. Item scores are then summed to create total scores for both effective and ineffective behaviors.

Recommendation

The State Statutes (Ref. Florida Meritorious Personnel Program, Section 231.533 of the State Statutes) require documented superior performance evaluation based upon scores from three observers to qualify for merit pay.

In using data from the FPMS for purposes of evaluation, it is recommended that the semi-interquartile range be used for setting standards. Examples of standards include:

Percentile Rank Score		Probability of Passing both	Percentile Rated Superior
Effective	Ineffective	Effective and Ineffective	
20%	20%	(.8 x .8 = .64)	36th
30%	30%	(.7 x .7 = .49)	51st
40%	40%	(.6 x .6 = .36)	64th
50%	50%	(.5 x .5 = .25)	75th

If, for example, the standard is set at 30% on the norm scale, then 49% of the teachers observed in the norming study would receive superior classroom performance evaluations. In setting this standard, other criteria used to qualify candidates for merit pay must be considered, i.e. experience, degrees, attendance, and subject area examination. Each of these criteria will screen out potential candidates. This implies that if the standard on the performance observation is set too high, few if any teachers will qualify for merit. Conversely, if the standard is set too low, then poorly performing teachers may qualify for merit. We suggest consideration of the 30th percentile as a cut score for both tests (positive and negative), which would define those individuals falling above the median on the FPMS who also qualify under all other criteria, as meritorious.

04/10/84
CPT:NS1-36

Section II

DEVELOPMENT OF NEW DOMAINS OF PERFORMANCE

During the 1982-83 academic year, the Domain documents were used by many school districts throughout the state of Florida. As the materials were used, the need to reorganize Domain 6 and to develop two new Domains was recognized.

Domain 6 was different from other Domains, for it included teacher behavior that occurred prior to teaching a class as well as in-class behavior. The behaviors an observer would expect to see during an observation of a teacher, who planned to evaluate his/her students, would likely include the administration of an examination, or perhaps, the preparation of students for subsequent administration of a test. What the observer would not expect to see would be the preparation of that test by the teacher. Domain 6 had a mix of concepts that had been gleaned from research on evaluation that covered both in- and out-of-class activities. For these reasons, it was difficult to design an instrument for this Domain. Consequently, it was felt that Domain 6 could be improved by separating the research on these two activities into two documents.

Development of Domain 6

This Domain has been revised more than any other Domain because the content did not fit together in the same way as other areas did. As mentioned above, some of the concepts were observable in the teacher's classroom performance, while other concepts were observable by interview and

examination of teacher's products. We divided the work into these two areas: (1) concepts observable in the classroom, and (2) concepts observable by interview and examination of the teacher's products. The first area became the revised edition of Domain 6.0, which was re-titled Testing: Student Preparation, Administration, Feedback. It was defined to include what the effective teacher does in the classroom prior to testing as well as during and after testing. The research for the second area, which includes the principles of test construction and the principles of rating and marking, will be included in a handbook on test development and evaluation.

From the survey of the literature, we identified three concepts comprising the revised Domain 6 together with the behavioral indicators relevant to each concept. These concepts and indicators were then defined in keeping with the research, and both positive and negative examples of each indicator were given.

The next step in the procedure was to formulate a succinct statement of the effect of the indicators on either student achievement or conduct, or both. Then followed a summary of the research evidence, including references. Where there were exceptions to the research findings, or research extending the findings to other populations, or research which provided alternative teacher performance, these were added.

This procedure allowed us to systematically survey the literature and at the same time organize the results into a form that proved to be useful in (1) instrument development, as described in a different section of this report, and in (2) preparation of instructional materials to help teachers remedy their deficiencies, as exemplified by the learning packages that are

being developed by members of the Florida Coalition.

All of the Domain documents are limited to generic knowledge and skills of teaching. By "generic," we mean the knowledge and skills that are not content specific. They do not directly pertain, for example, to the preparation of mathematics students to do arithmetical procedures, except as part of the test content, but rather they include knowledge and skills that cut across more than one subject and more than one grade. For example, the skill of instructing students to utilize the characteristics and format of a test, such as reading test directions, first marking the test items they know and omitting the hard items, checking over all the answers when finished, etc., are used from subject to subject and grade to grade, sometimes with a degree of modification. (See Appendix Section II, A for Domain 6.0.)

Instrument Development

Two formative instruments for the documentation of teacher performance were developed: a product checklist to be used with the Handbook and a performance observation schedule for the revised Domain 6. The product checklist has two categories: test construction, and the rating and marking practices of effective teachers. This checklist requires the consultant to examine products, such as a teacher's tests, and to interview the teacher.

Trainers in the field have continued to use the product checklist in its present form, although it is tentative. The Handbook Committee on test development and evaluation has not reported on its work as of yet, so we do not know if the product checklist is comprehensive or if it will need to be

revised by the committee.

The Domain 6 instrument--the performance observation schedule--contains three concepts: preparation for testing, test administration, and formative feedback. Findings from the literature indicate that specific teacher behaviors, such as providing frequent testing and feedback, may be effective in increasing student achievement. Effective teachers may reduce student fear of examinations, improve student test-taking skills, reduce cheating, and influence student attitude regarding the evaluation of their work.

While some teacher behavior has positive effects, other specific behavior shows a negative relationship to student learning. The instrument format, therefore, provides for the coding of both effective and ineffective teacher behavior, in the same way as the formative instruments for Domains 2-5 do.

This instrument, like the other instruments in the Florida Performance Measurement System, avoids the common error found in many evaluation systems of combining observation and evaluation into a single rating scale. This error forces an observer to perform two distinct operations at one time: (a) to observe and record accurately, i.e., to measure, and (b) to judge the quality of the performance. This instrument was developed for the first purpose only. The observer records the data, and they are analyzed and interpreted separately. (See Domain 6.0 Formative Instrument in Appendix Section II, B.)

Development of Domain 7

The research findings in Domains 1.0 through 6.0 document 124 specific teacher behaviors that have been shown to be positively related to increased student achievement and improved classroom conduct. The next two domains are somewhat different. Domain 7.0 focuses on the teacher's behavior in an individualized setting, and Domain 8.0 deals with the special problem of helping the teacher. The latter documents research on how a consultant should communicate classroom observation data to the teacher.

From the beginning, the research team was aware that the teacher's job was multi-faceted; there were responsibilities he/she must perform beyond the classroom hours of contact with the students. Preparation for classroom instruction; conferences with administrators, other teachers, and parents; staffings; and faculty meetings, is to name a few of the other activities that can be part of a teacher's day. All of these were important and were slated to be considered in the FPMS. However, the major focus of the system has been on the primary function of the teacher--classroom instruction. The materials developed thus far from Domains 1-5 have all been concerned with group instruction.

The use of the FPMS for the observations of teachers in the Beginning Teacher Program has highlighted the need to look at teaching in a second setting--the one-on-one session. Although research supports instruction to the total group as an efficient mode of delivery for the usual pattern of classwork, there are some areas of instruction which are by their nature individualized, and generally require the teacher to work one-on-one with the student. Most of these areas are not in the usual classroom and the

teachers are not classroom teachers but specialized instructional personnel such as speech therapists, school psychologists, counselors, etc. Because the Beginning Teacher Program includes a large number of these specialized people, the need to develop a specialized instrument requires no defense.

The first approach to the development of this special use of the Domains was to find what was appropriately used in the field. We wrote some examples specific to the special area field. For example, punctuality is important in group instruction and in individual sessions as well. An example of effective punctuality in a one-on-one session written for one group of specialists, School Library Media Specialists, follows:

> When John came to the media conference room, the librarian was ready to begin. She had materials organized, and began the session promptly by greeting John and inviting him to sit next to her at the conference table. She then began the discussion by stating, "Mr. McQuire told me that you would like to explore some possible careers in woodworking. Can we begin by you telling me more about your interest in this?"

The second step in the development of Domain 7 was to compile the concepts and indicators with their definitions in a draft document. This draft contained concepts borrowed from Domains 3 and 5. In addition, the content from Domain 8 on Conferencing, discussed below, was included. This Domain is tentative in its present form, but it is a compilation of research findings that seem to have application for the special instructional personnel who work with students and others in an individualized setting.

We are working cooperatively with the special area personnel in an attempt to develop appropriate materials. A draft of Domain 7.0, together with some examples to illustrate application of the concepts and indicators, was presented to representatives of special areas in the State Department of

Education. The procedure suggested was for each special area representative to first look carefully at the concepts and indicators that were supported by research and then to develop positive and negative examples for their own area.

Next, if additional generic concepts can be documented by research, these items were to be recommended for inclusion. The area teams were cautioned to avoid advocatory or textbook materials, and to select low inference items.

In February, we met with a selected group from special areas who have been developing the materials for this Domain. They reported on their progress. For this year, they planned to use the FPMS materials in the same format as everyone else. For the future, they will develop materials for parts of their jobs that are not group instruction. It will take time to assemble a literature review for one-on-one instruction, consultation, management skills, and other job responsibilities. Assessment of some of these roles may be by product examination or interview rather than by observation of behavior. The selection of library materials by the librarian is an example of a nonobservable item to be included in an evaluation of the librarian's performance.

Domain 7.0 continues to be a draft document. It will need revision by the specialists, and when the development is completed and an instrument or instruments are formulated, it will need to be validated for the one-on-one setting. (See Appendix Section II, C for Domain 7.0 and some examples.)

Development of Domain 8

Domain 8.0 has undergone a metamorphosis somewhat like Domain 6.0. Although Domain 6.0 was divided because some of the concepts were observable and some were not, Domain 8.0 was divided because some concepts were related to communicating the observed data to another person and some were not. The Domain was originally titled "Consultation." It was proposed to document the skills that the principal, supervisor or peer teacher would exhibit when that person met with a teacher before or after an observation as well as during the intervening conferences.

The term "consultant" was selected because it typically was used in research literature wherever someone served in a helping capacity in reference to problems of planning, instruction, evaluation and student conduct. The research in this Domain was gathered with the idea of developing a complete digest of conferencing skills. However, it was learned that the Council on Educational Management was developing a module on Conferencing Skills for Administrators so our focus was narrowed. We decided to limit the Domain on Consultation to skills needed for sharing information about data collected during an observation. Domain 8.0 was consequently retitled, "Communication of Observation Data."

The concepts covered in this specialized domain include the development of rapport, interpersonal skills, problem formulation, and development of plans. Other research on consultation that had been reviewed was included in Appendices A and B of the Domain. It dealt with the question of whether or not consulting is effective in changing teacher performance and whether it is better to use a direct or a collaborative approach to consultation.

This Domain, with appendices and an extensive bibliography, was presented to the Committee on Conferencing of the Council on Educational Management. We expressed the hope that this Committee would incorporate the research and the concepts and indicators developed from the research in their module on conferencing skills for school administrators. We did not continue to develop the consultation materials for this Domain. Specifically, we did not develop an observation instrument to measure the performance of administrators when they actually engage in a conference with a teacher. This will be a next step, but we need to see what other conference skills are extracted from the research by the Committee from the Council on Educational Management. This work needs to be a joint effort so that there is no further duplication of the work involved.

One of the reasons for the overlap is that the skills of communicating data to others are not exclusively in the arena of one type of administrator. The Council on Educational Management is concerned with the training of school administrators, and this module on conferencing skills is only one of nine modules they are preparing for the training of principals. Our work has focused on the teacher, particularly the beginning teacher. As we worked with the trainers in the Beginning Teacher Program the past two years, the need for specific information about how to communicate effectively with the beginner was expressed again and again. The trainers were called upon to help principals and many other persons such as supervisors, staff development workers, and peer teachers who want assistance in how to use observation data effectively after collecting them. Training in communicating observation data needs to be delivered to anyone providing help to a teacher

whose need for improvement is decided on the basis of data derived from observation of classroom performance. (See Appendix Section II, D for Domain 8.0.)

As work on this special area continues next steps will include, in addition to the development of an instrument for this Domain, the validation procedure to test and revise the instrument and the identification and development of materials for the improvement of all of the consultant skills that were identified.

To summarize briefly, the addition of new research will constantly require an updating of the materials in the FPMS, but as for now, the present state of all the Domains except 7.0 is no longer tentative. These Domains are the most extensive collection of research findings on teacher effectiveness available.

Recommendations

There is need for information to improve the teaching of the forms of knowledge--concepts, laws, values, rules, etc. We are beginning to collect research having to do with the higher mental processes. The next step should be to identify concepts and indicators of effective teaching that engage student thought in complex high-level cognitive processes and abstract forms of knowledge. In addition, the research on teaching of content specific knowledge and skills--those required for effective teaching of mathematics, science, etc.--must be assembled and digested. These further studies will yield important new Domains of teacher effectiveness.

Section III

A PROPOSED FEEDBACK SYSTEM

The public schools of Florida employed approximately 4500 beginning teachers in the academic year 1983-84, or about 5 percent of the total teaching personnel. Approximately 2000, or 45 percent, are graduates of Florida institutions of higher learning. The remaining 55 percent were recruited from other states.

The 1983 survey of follow-up practices of Florida colleges and departments of education (see Teacher Evaluation Project, Final Report for 1982-83) shows that there is no uniform system of data collection concerning the quality of the graduates of either in-state or out-of-state institutions. Only 15 of the 30 Florida institutions preparing teachers reported a follow-up study of their graduates. Of these, eleven reported that they made annual surveys and four reported only periodic surveys. None collected objective data about the classroom performance of their graduates.

Eleven of the 15 follow-up studies tapped the opinions of graduates about the quality of instruction in pedagogy and fourteen surveyed employment status. It is interesting that only seven schools, less than half, reported that the findings of their surveys were used either to evaluate or improve their programs of teacher education.

Purpose of Feedback Study

The purpose of the present study is to suggest a systematic plan for collecting and analyzing data about beginning teachers to be used in

--developing and monitoring support team programs to improve the performance of beginning teachers during their initial year of service,

--setting minimum performance levels for certification of beginning teachers,

--developing programs in teacher education colleges and departments,

--developing programs for inservice staff development for beginning teachers,

--identifying problems of the Beginning Teacher Program and designing research relevant to them,

--providing information to school districts about the quality of out-of-state beginning teachers and identification of institutions from which they came.

Feedback Components of Teacher Education

An effective system must be broad enough to cover all the major facets of a teacher's formal preparation. At least three components must be included. The first is academic preparation typically provided by non-education departments such as those in Liberal Arts Colleges. This component constitutes 70 to 85 percent of the total hours of a teacher's education. The second component consists of professional studies. These studies are taken in a College or Department of Education and constitute 15 to 30 percent of the hours required for graduation, including an internship of approximately ten hours. The third component consists of the first year of teaching under the supervision of a support team. This team identifies behaviors on which the teacher needs improvement and provides assistance

relevant to these behaviors. Future improvement of teacher education is dependent upon knowing the effectiveness of each of these components.

In order to evaluate the first component--academic subject matter preparation--subject matter examinations should be required by the State for admission to beginning teacher status, and should also be used to determine the weak spots in academic preparation provided by Florida colleges and universities.

Development of appropriate examinations will require an analysis of the public school curriculum at all levels to determine the scope and depth of knowledge required of an effective teacher. It will also require analyses of the logical operations performed in presenting and explicating knowledge at different levels of instruction and familiarity with the forms of knowledge found in the school curriculum. Only academic examinations that cover these dimensions can yield data appropriate to the task of improving the subject matter preparation of teachers.

The second component--preparation in pedagogical knowledge--is assessed by the present professional knowledge examination, passage of which is required for beginning teacher status. This examination, although valuable and necessary, tells little about how well the preservice program has prepared the teacher for performance in the classroom. However, it indicates how well the concepts and principles of pedagogy have been mastered.

The third component--Beginning Teacher Program--is designed, among other things, to assist the beginner in the effort to expand and improve his/her repertoire of instructional skills. Its aim is to help the beginner adjust to the job situation and to improve classroom performance.

The Proposed Feedback System

The feedback system consists of procedures for collecting and analyzing data about the performance of beginning teachers, their preparation, assignments, and opinions about their preservice preparation. It also includes an agency for coordinating and processing the data and communicating the results of data analysis to the schools and colleges.

The establishment, implementation, and maintenance of a state-wide feedback system places heavy reliance on the administrative and supervisory personnel of the schools and the State Department of Education. The accuracy of data collection must be safeguarded; the procedures and techniques of data analysis must be carefully selected and designed; and strict time lines must be adhered to for completion of each phase of the total feedback operation.

As indicated above, two sets of data are to be assembled: data about the prepartion, assignment, and opinions of the teacher and data about classroom performance. Two approaches to the collection of these data have been considered. The first places the major responsibility upon the principal and his staff. It requires the principal, or designee, to sum the tallies of each indicator on a summative instrument for each beginning teacher, and to transfer these summations to an OPSCAN form in accordance with the following instructions:

Step 1. For each teacher, transfer the total tallies for each effective indicator from the instrument to a specifed OPSCAN sheet. (See Appendix Section III, A for an example of an OPSCAN form.)

Step 2. Repeat the procedure in Step 1 for each ineffective indicator.

In addition, this approach requires each beginning teacher to respond to a questionnaire (see Appendix Section III, B) and to transfer his/her responses to the same OPSCAN form as that specified for the reporting of performance tallies.

The second approach is designed to collect the same information as the first approach but without placing as much responsibility for assembling the data upon the principal. It requires that the principal provide copies of the tallied summative instrument, both initial and terminal observations, for each beginning teacher to a coordinating agency. The tallies for each teacher will be summed and transferred by key-punching to a computer, programmed to yield the desired information.

The teacher is to record his/her responses to the Beginning Teacher Questionnaire on the OPSCAN form just as required in the first approach. These forms are collected and transmitted by the principal to a coordinating agency along with copies of the summative instrument tallies.

We have field tested aspects of these two approaches in Hillsborough and Pinellas School Districts. These tests led to correction of ambiguities, hazy expressions, and substantive errors in the Questionnaire. They also provided information about time requirements. Both effective and ineffective tallies for the summative observation of a single teacher can be summed and transferred to an OPSCAN form in ten minutes. A school with six beginners, for instance, would require approximately one hour of the principal's time to prepare the OPSCAN forms for either the initial or terminal observation. If the second approach were followed, the principal would simply make duplicate copies of the summative

observations and forward them to the coordinating agency.

The time required for a teacher to complete the Beginning Teacher Questionnaire ranges from twenty to forty-five minutes, depending upon the individual. However, this Questionnaire can be completed either at school or home, depending upon how the principal handles the task. In either case, the time required of the principal is negligible.

It is apparent from the foregoing discussion that time requirement is not a prime consideration in deciding which of these two approaches is preferable. The telling factor is not time but accuracy. Which approach is likely to entail the fewest errors? If the principle that errors occur in direct proportion to the number of persons handling the data is valid, the second approach is preferable. This approach can require as few as a single key-puncher, one who can readily become acquainted with the sort of questions that invariably arise when responses and tallies are being processed, and can work out consistent answers and interpretations. In sharp contrast is the first approach where two or three thousand persons--principals and their staffs--will be answering the same questions independently. This procedure would appear to bode wide variation and multiplicity of errors.

The coordinating agency will provide information to the districts and teacher education institutions about the strengths and weaknesses of beginning teachers. It is not instructive to report the frequency with which a particular teacher performs each indicator on the observation instrument. For a report to be meaningful there must be a standard of performance against which a teacher, or the teachers who are products of a par-

ticular college, can be compared. Such a standard typically takes the form of a norm or cut-off point below which a teacher is said to be deficient or a college program is judged to be inadequate if a preponderance of its students do not measure up to the norm.

The research being conducted under the auspicies of the Florida Coalition, described earlier in this report, provides statistical information by which to decide performance norms for Florida beginning teachers. The coordinating agency, on the basis of this research, can report to each teacher education institution where its products stand with respect to the norms of beginning teachers, how its products compare to the products of other Florida Schools and out-of-state schools, and the domains of performance where its graduates are strong and where they are weak.

Moreover, the coordinating agency can report the opinions of the graduates of each institution about the quality of the preparation in the various performance domains. While teacher satisfaction with their education is not necessarily a measure of the quality of either academic or pedagogical preparation, it does evoke questions about programs of instruction when the incidence of dissatisfaction is high.. Furthermore, satisfaction - dissatisfaction ratings are more likely to be valid when teachers can compare their preparation with job requirements than when they are inexperienced students in a preservice program.

The Coordinating Agency and Its Tasks

The collection of data pertaining to the classroom performance of hundreds of beginning teachers each year, not to mention the likely increase

in the number of beginners from year to year, is an operation of considerable magnitude. The scope of the feedback operation and its potential for improving teacher education at all levels strongly suggests that the coordinating agency be adequately funded and located in the Office of Teacher Education, Inservice Staff Development and Certification in the State Department of Education.

Among other tasks, the feedback operation will require that

--the beginning teacher questionnaire and summative instrument be supplied to each district at the beginning of the school year,

--the questionnaires and summative observations be completed following the first performance observation and returned in proper order to the coordinating agency,

--arrangements be made for the collation and analysis of the data from the first and second observation by the coordinating agency,

--the coordinating agency inform each Florida College or Department of Education about the performance of its products, indicating how its graduates stand with respect to performance norms for beginning teachers and the effective and ineffective indicators of a summative instrument,

--the districts forward copies of all terminal observations or OPSCAN forms to the Coordinating Agency at or near the end of the school year, and

--the coordinating agency inform each district about the year-end performance of its beginning teachers, indicating gains of each beginner with respect to performance norms and the effective and ineffective indicators of the summative instrument together with such district comparisons as is deemed desirable.

Recommendations

1. In order to safeguard the Beginning Teacher Program and to stimulate improvements in the preservice programs of Florida colleges and departments of education, it is recommended that a feedback system along the lines of the second approach set forth in the foregoing analysis be established by legislative action, adequately financed, and located in the State Department of Education.

2. Since the Office of Teacher Education, Certification and Inservice Staff Development is closely associated with all aspects of teacher education, it is recommended that the feedback operation be centered in a coordinating agency in that Office.

3. Given the fact that there is now no satisfactory set of examinations covering the content of the various teaching fields and the additional fact that knowledge of the subject matter of instruction is essential to effective teaching and therefore to proper certification, it is recommended that the legislature extend S231.0883 of the Florida Statutes to cover beginning teachers, and to take such further steps as it deems necessary to develop appropriate teacher examinations in all subject matter fields of instruction represented in the curriculum of Florida schools.

04/10/84
CPT:SSIX1-12

Section IV

DEVELOPMENT OF THE FPMS COGNITIVE EXAMINATION

Observation instruments of the Florida Performance Measurement System were developed from research on teacher effectiveness. Findings from the studies indicate that certain teacher behaviors can be effective in increasing student achievement while others show a negative relationship to achievement.

In the FPMS, the observer does not make judgments as to the quality of the teacher's performance; instead, he/she only records the teacher's behavior. Analysis of the recorded data is a separate function that occurs at a later time. While there is little doubt that persons who analyze observation data must know the research literature, there is some question as to whether or not the observer must also know the research base of the indicators on the observation instrument.

Why Knowledge of Research Findings Is Necessary

The need-to-know can be illustrated in several ways. First, the observers may tally a teacher behavior in the wrong category because they do not understand the terminology in the Domain documents. For example, a teacher merely referred to a previous lesson(s) by stating, "Yesterday we studied simple sentences. Today we will study compound sentences." The observer tallied this as "review." This indicated a lack of congruence with the research definition which states, "teacher performance that either rehearses the main points of a previously discussed topic, problem, unit or

lesson(s) or that directly involves the class in the rehearsal." This is a grievous error. We learn in Domain 3 that "if reviews are conducted at the end of the lesson and at weekly intervals (or occasionally longer ones), then retention as well as learning will increased." Furthermore, the observer has given the teacher credit for a review that did not occur. When the data are analyzed, the teacher will not receive the appropriate feedback that might help him/her improve this important part of lesson delivery.

A second reason the observer needs to know the research literature involves the comprehensiveness of the indicators on the summative instrument. Each indicator represents a body of knowledge which the observer should know in order to give the teacher credit for performance. For example, in Domain 2 there are a number of ways a teacher might stop misconduct. The obvious procedure is a direct desist, e.g., "John, stop talking." However, unless an observer is familiar with the research, he/she might fail to tally "stops misconduct" when there is a disruption and a teacher (1) suggests an alternative behavior to John, e.g., "John, will you please read the next problem?" or (2) alerts the class (indirectly silencing John) by saying, "Each of you should listen to my questions for you may be called on next," or (3) praises another child for compliance, e.g., "Kay, you have been doing an excellent job of listening," or (4) reminds the class of a rule, e.g., "Class, what is our rule about listening?" These less obvious ways of controlling John's behavior may be equally or even more effective because John does not receive the attention he covets.

Moreover, it is important for the observer to understand the research in order to understand why some behavior is designated as ineffective,

although it may sometimes be properly used by the teacher, given the context or purpose of the lesson. A case in point is "poses non-academic questions." We know from research that asking opinion questions does not directly improve achievement; therefore, it is not named an effective indicator. However, a teacher wants to motivate students by getting them to relate their personal experiences to a lesson, such as asking "How many of you have a pet?" or "What do you think would be a good name for the blind boy's dog?" Knowing the literature helps the observer accept the location of this indicator and others on the ineffective side of the FPMS summative instrument.

Experience from working with trainers suggests some of the difficulties in using the FPMS summative instrument without an adequate knowledge of the literature. Much of the difficulty in the use of an instrument is in learning the vocabulary that is unique to the instrument. Praise is an example. Observers must learn to recognize specific academic praise as contrasted to general, non-specific praise. This is the key to the important follow-up procedure of helping teachers improve. Whether the data are used to help the individual teacher, or analyzed to assess the inservice needs of a faculty, or compiled to provide feedback to preservice colleges and departments of education about the needs of their graduates, the first step is to assure the competence of personnel who will collect the data. That competence must include understanding of the research literature upon which the data collection instrument is based. To this end, a comprehensive training program must be pursued, followed by an equally comprehensive testing and evaluation procedure.

Development of the Cognitive Examination

The first cognitive examination (Form A) was developed to assess the research knowledge of the corps of FPMS trainers. The 71 item test on Domains 2-6 was administered to all trainers at the close of a three-day training workshop. Observation tests were also administered to all trainers using two videotaped lessons. A revision of Form A based on an item analysis was underway at the time it was announced that the FPMS would be used for the observation of teachers in the State's Associate Master Teacher Program.

To expand the number of cognitive test items, the developers contracted a group of persons knowledgeable of the FPMS (see Test Item Writers in Appendix Section IV, A) who were available to begin work immediately. This was a crucial requirement because the plan set forth by the Department of Education required a complete test for the assessment of observers to be administered on January 23-26, 1984, in 12 assessment sites throughout the State. The group of item writers consisted of an elementary school principal, a high school principal, several staff development people, a resource teacher, university faculty members, staff members who had been involved in the development of the Domains, and those who were trained in the delivery of the FPMS.

Item writers began developing test items during the 1983 Christmas holidays. Each writer prepared 15-20 items from a pre-assigned Domain. The team of item writers met at University of South Florida on January 13, 1984, to review the test items that had been developed, and to prepare a Form B of the cognitive examination. Individual items were reviewed by teams for con-

tent, test format, and correctness. Test item composition was decided by the number of concepts in each Domain. The three best items, as judged by the team of writers, were selected to represent each concept. For example, Domain 2 has eight concepts. At three items per concept, there were 24 items chosen to represent Domain 2 on Test Form B.

Following this meeting, the developers reviewed the items for content and comprehensiveness. The items were then edited for consistent format, spelling, grammar, and punctuation. The answers were checked to assure freedom from an answer pattern.

Administration of the Cognitive Examination

By January 18 a final copy was typed, proofed and duplicated. Thirty-one hundred copies were numbered and shipped to 12 assessment centers designated by the Department of Education. Transoptic answer sheets, directions for their use, an inventory sheet, and a test security form were included with the tests.

A major concern of the test administration has been security. Close cooperation with the Department of Education on this matter was maintained. The Department of Education personnel selected twelve FPMS training consultants whom they prepared to administer the test. The Department of Education also prepared a manual of instructions to insure consistency and security in the testing operation. In addition, the verification of test security that was signed by the test administrator and the test monitor required them to stay in the test room and keep the tests in sight at all times. All of the materials were returned in one box and checked to verify

the security of the test copies.

Development of a Test Item Bank

To continue to improve the quality of FPMS observations, an examination on research knowledge as well as tests of observer accuracy will be needed. Therefore, 500 additional items were prepared so that different test forms can be assembled as needed. The item writers established February 29, 1984, as the deadline for the completion of additional items from each Domain, including Domain 1.0 which had not been a part of Test Forms A and B.

The following list of item specifications was suggested and adopted by the test item writers for the review of items and content analysis.

1. All items should be SECURED at all times.
2. All items should be bias-free.
3. Item format should be multiple choice or matching.
    a. Item stems should be positive when possible.
    b. Answer options should be limited to four (A; B; C; D).
    c. Answer options should be alternative statements, not "all of the above," "none of the above," etc.
4. Item content should be drawn from the research or the concepts and indicators in the Domain document.
5. Items should represent all of the concepts within the Domain.

Table 1 was used to inventory the number of items that have been prepared in each Domain.

TABLE 1

Inventory

List number of items by Domains.

Domain	Revised Test A Items	Test B Items	New Items	Domain Totals
1	0	0	76	76
2	12	24	61	97
3	8	16	88	112
4	24	13	52	88
5	8	15	57	80
6	9	7	30	46
Totals	61	75	364	500

The analysis of individual test items included naming the origin of each acceptable test item, i.e., whether the item content was taken directly from the research literature in the Domain, or whether the question was derived from the definition of the concept or some other part of the Domain narrative, such as the examples or the principles which were formulated from the research. This designation was documented by citing a page reference in the Domain document. The representativeness of the item was noted by giving the concept number and/or indicator number that was directly related to the test question. This step was important to assure the team that all of the concepts in the Domain were covered by the questions. This information will be needed to develop tests that have a balance of items, instead of loading items on one or more concepts in a Domain.

The final step was an analysis of the level of knowledge that each item

required the test-taker to use. The distinction between recognition or recall items and items which measured interpretation or application was listed for items in some of the Domains. Because there is research to suggest that test-takers use varied cognitive skills to solve the same problems, this analysis was deemed unnecessary and it was decided to provide for variety in item format (Nuthall, 1982).

The team of item writers reviewed each item for Domains 1.0, 3.0, 5.0 and 6.0 and suggested revisions where the item did not meet specifications, or when it was judged unacceptable for any reason. The content validity of the items is now thought to be satisfactory.

Altogether there are now 500 items that could be used to compose tests on Domains 1.0 through 6.0. The present Study Guide for the Cognitive Examination indicates the need to review Domains 2.0 through 6.0 inclusively for Test Form A and Test Form B. (For the study guide, see Appendix Section IV, B.) As various test forms are devised in the future, revisions in the FPMS Study Guide should be made to reflect the content to be covered.

It is important to keep test items secure. To this end, each item writer signed a verification of Test Security Form.

Recommendations

1. We recommend that if and when the Florida Performance Measurement System is adopted by the Florida State Board of Education for use in Florida school districts, the Department of Education assume the responsibility for test construction and test administration of the FPMS. This would help to assure the test security that is needed in the process of

evaluating and certifying both trainers and observers, whether they are university or public and private school personnel.

2. We also recommend that the test items developed this year, and any new items, be subjected to further validity analysis.

04/10/84
CPT:SSEVEN1-7

Section V

PROPOSED REVISION OF FLORIDA GENERIC COMPETENCIES

The Florida Generic Competencies constitute the essential knowledge that teachers must possess to be considered qualified professionals. These competencies, formulated by representatives of the teaching profession, were approved by the State Board of Education and thereby bear the force of law. They are the criteria which superintendents must declare a beginning teacher has satisfied in order to be recommended for certification.

What Are Generic Competencies?

The term "competency" is defined, after Medley (1982), as a single item of knowledge or skill relevant to the successful practice of teaching. The ability to give specific academic praise is a competency. In this sense, competency refers to what the teacher does, or is able to do, and not to the effects of the doing.

The expression "teacher competence," or "competence of a teacher," designates the total stock of competencies a teacher possesses. A teacher can have varying degrees of competence depending upon the number of items of knowledge or skill comprising his or her repertoire. However, the effectiveness of a teacher is not the same as his or her competence. Competence is what s/he knows or can do. In sharp contrast, effectiveness is the result of teacher performance on student learning; performance is on-the-job behavior of the teacher; it can be more or less effective depending on its properties.

Some competencies are common to all teachers; they cut across all grades, subjects of instruction, and types of students. These are called generic competencies. However, there are also specialized competencies -- those which are unique to subjects of instruction or to particular positions such as media specialists, librarians, and counselors. These are not treated in this report. Committees are now assembling and organizing the competencies of these specialties. (See Section II of this Report.)

Origin of the Florida Generic Competencies

The Florida Generic Competencies were developed under the auspices of the Florida Council on Teacher Education to serve primarily as criteria for certification of educational personnel and as a program focus for Florida institutions of teacher education. The Florida Council on Teacher Education, created by Section 5 of Chapter 231.10 of the Florida Statutes, was appointed by the State Board of Education to advise the Commissioner of Education on all questions of teacher education and certification.

Early in 1975 the Commissioner of Education appointed a task force to review teacher training, certification, and the various state councils having to do with teacher education, certification, and professional practices. This task force recommended that a study group be created to propose indicators of desirable performance. The Commissioner of Education then asked the Council on Teacher Education to design and conduct studies in response to the recommendation of the task force.

The Council on Teacher Education began the competency studies early in 1975. These studies consisted of two kinds. The first was designed to

identify competencies to be required of all certified personnel, and the second was designed to identify those competencies considered to be unique to specialized areas of certification. The second set of studies was terminated when the Council on Teacher Education was displaced by the Education Standards Commission in 1981.

The procedure followed in the first set of studies was to make a thorough search of pedagogical literature to ascertain those competencies that were vouched for by authoritative sources. Forty-eight competencies resulted from this review of the literature and they were subsequently submitted for ranking to a random sample of certified educational personnel in August 1975. A high degree of consensus was obtained among judges for 23 of the 48 competencies.

These 23 generic competencies were recommended to the Commissioner of Education for incorporation into State Board of Education rules governing teacher education and certification, and were subsequently incorporated into the rules in March 1978. Then in 1982 an additional competency on instructional needs of exceptional students was adopted, making a total of 24. Along the way, competencies 6 and 19 were combined, reducing the total to 23.

All of the 23 competencies, except for 2, 3, 4, 5, and 6/19, are supported by research, and these exceptions have face validity. (See Appendix Section V, B for original list of competencies.)

Suggested New Competencies

In the brief time since the competencies were formulated, research on teaching has substantially expanded pedagogical knowledge. This knowledge has now been assembled and organized by a group working under the auspices of the Coalition for the Development of the Florida Performance Measurement System. New competencies have emerged from this knowledge and they are herein proposed as additions to the 23 competencies. These additional competencies are as follows:

1. Demonstrate ability to stimulate and direct student thinking and to check student comprehension by questioning techniques.
2. Demonstrate ability to provide practice to promote learning and retention.
3. Demonstrate ability to treat student talk in ways that encourage student participation and that maintain academic focus.
4. Demonstrate ability to use feedback procedures that give information to the student about the appropriateness of his/her response.
5. Demonstrate ability to conduct review of subject matter.
6. Demonstrate ability to use class time efficiently.
7. Demonstrate ability to present forms of knowledge such as concepts, laws, and rules.
8. Demonstrate ability to control the quality of vocal expression.
9. Demonstrate ability to use body language to express interest, excitement, joy, and positive personal relations.
10. Demonstrate ability to give examinations in a manner to minimize

anxiety and cheating and to provide appropriate feedback on test performance.

Development of Subcompetencies

These new competencies, like those in the original list, are not, strictly speaking, competencies, but rather clusters of competencies. For example, competency number 5 above, "Demonstrate ability to conduct review of subject matter" is composed of at least three competencies: a) ability to rehearse the previous lesson(s) (daily, weekly, monthly) or to involve students in doing so; b) ability to recap the significant points of a preceding discussion before moving to a new aspect of a topic or problem; and c) ability to restate, or to induce students to restate, the content of a discussion at the end of a lesson.

In developing the original list of competencies, it was early recognized that the competencies were too general for use in either test construction or development of classroom observation schedules. A committee was appointed to analyze each competency into subcompetencies or subskills, as they were called. The subskills, however, were pitched at a high inference level and consequently provided little more information than the competencies themselves. For example, the subcompetency "Demonstrate effective techniques and strategies for managing student behavior" does not specify these techniques and strategies. It does not tell what to do to execute the main competency; namely, "Identify causes of classroom misbehavior and employ a technique(s) for correcting it." The research findings on effective teaching, assembled under the auspices of the Florida

Coalition for the Development of the Performance Measurement System, however, provide low inference subcompetencies appropriate to most competencies of the original list and to all the proposed additions.

In order to conduct valid and reliable classroom observations the competencies must be analyzed into low inference teacher behaviors. In keeping with this requirement, the revision herein proposed replaces the present subcompetencies with low inference subcompetencies derived from research literature, except those for competencies 2, 3, 4, 5, 6/19, and 24 which are not conveniently evaluated by observation.

Classification by Modes of Evaluation

Since 1980, the competencies possessed by prospective teachers who seek a Florida certificate have been evaluated by objective examinations covering almost all of the 23 competencies. These examinations tell whether or not a prospective teacher has a minimum level of competence in academic skills, ability to read and interpret pedagogical literature, and ability to make valid decisions about what to do in educational situations. These tests mark a distinct advancement in the cognitive evaluation of teachers for certification and should be continued with increased zeal for their improvement. Perhaps steps should be taken to explore ways of more closely simulating actual classroom situations through the use of media such as videodiscs.

The competencies possessed by teachers can also be evaluated in other ways. Some can more conveniently be evaluated by think-aloud procedures, interviews, or analysis of products. Still others can be evaluated more

effectively by objective classroom observation.

In the practical situation of deciding which competencies are satisfied by a teacher, it is efficient to know how data on them can best be secured. For this reason, the competencies are reordered and classified into the following groups:

Group 1. Competencies most appropriately evaluated by cognitive tests.

Group 2. Competencies most effectively evaluated by think-aloud or interview procedures or by analysis of products.

Group 3. Competencies most effectively evaluated by observation of teacher behavior in the classroom.

(For the revised list of competencies and subcompetencies classified by modes of evaluation see Appendix Section V, B.)

Summary of Proposed Changes

The proposed changes are set forth in the following summary:

1. Competency 1 of the original list has been placed in Group 3 and Competency 24 has been placed in Group 1, because Competency 1 is observable and number 24 can best be evaluated by cognitive tests.
2. Ten competencies have been added to the list of 23, making a total of 33 competencies.
3. The original subcompetencies have been replaced by low inference subcompetencies except for those in Group 1. The latter are unchanged because high inference competencies are not an impediment to test construction.

4. The 33 competencies have been classified into three groups according to modes of evaluation.

Recommendations

In order for the foregoing revisions of the Florida Generic Competencies to be considered for adoption it is recommended that:

1. The proposed revision be reviewed by the Office of Teacher Education, Staff Development, and Certification.
2. If approved, it is recommended that the proposal be transmitted to the Education Standards Commission with the suggestion that it be submitted to a representative sample of the profession for its consideration and recommendation.
3. If approved by the Education Standards Commission, it is further recommended that the proposed revision be transmitted to the State Board of Education with recommendation that it be incorporated in the rules having to do with certification and teacher evaluation.

04/10/84
CPT:SNINE1-6

BIBLIOGRAPHY

Blair, R. Clifford and James J, Higgens. Treatment of hypotheses under various regression/coding methods for non-orthogonal factorial designs. Educational and Psychological Measurement. 1978.

Blom, G. Statistical Estimates and Transformed Beta Variables. New York: Wiley and Sons, 1958.

Domains: Knowledge Base of the Florida Performance Measurement System, Florida Coalition for the Development of a Performance Measurement System, Office of Teacher Education, Certification, and Inservice Staff Development, Tallahassee, Florida, 1983.

Kirk, Roger E. Experimental Design: Procedures for the Behavioral Sciences. Belmont, California: Wadsworth, 1968.

McDonald, Frederick J. A Report on the Results of Phase II of the Beginning Teacher Evaluation Study: The Effects of Teaching Performances on Student Learning. Paper presented at the annual meeting of the American Educational Research Association. San Francisco, California, April 1976.

Medley, D. M. Teacher Competency Testing and the Teacher Educator, Washington D. C.: American Association of Colleges for the Teacher Education, 1982.

Micceri, Thedore. A Technical Report on the Reliability Study for the Florida Beginning Teacher Evaluation Summative Instrument. University of South Florida, Tampa, Florida: 1979.

Nuthall, G. and Lee, A. A. Measuring and Understanding the Way Children Learn in Class. Technical Report: Teaching Research Project, Education Department, University of Canterbury, Christchurch, New Zealand, 1982.

SAS Institute Inc. SAS User's Guide 1979 Edition. Cary, N. C.: SAS Institute, 1979.

Teacher Evaluation Project, A Study of Measurement and Training Components Specified in the Management Training Act, Final Report for 1982-1983. Florida Coalition for the Development of a Performance Measurement System, Office of Teacher Education, Certification, and Inservice Staff Development, Tallahassee, Florida, 1983.

Appendices Section I, A-G

Norming Study

CONTENTS

Appendix		page
A.	INSTRUMENTS AND RESULTANT VARIABLES	1
	The FPMS Summative Observation Instrument	1
	The FPMS Frame Factor Instrument	2
	Nature and Content of Instruments	3
	The FPMS Summative Observation Instrument	3
	Instruction Format Instrument	3
	Frame Factor Instrument	4
	Variables	5
B.	SOCIO-DEMOGRAPHIC, CLASSROOM AND LESSON VARIABLES	7
C.	STANDARDIZING OBSERVATION SCORES	27
D.	QUARTILE BASED SCORING METHOD	31
	Techniques Applied in Developing Scale	32
	Notes	32
E.	TOTAL SCORES COMPARED ACROSS FRAME FACTORS	34
	Introduction	34
F.	PROCEDURE USED IN DEVELOPING ADJUSTMENTS	43
	Basis and Scores Used in Adjustment	44
G.	RESULTS OF POSITIVE SCORES	45
	Results of Regression Analysis	45
	General Results of Anovas	45
	Oneways	45
	Results following Adjustments for Instructional Method	46

LIST OF TABLES

Table		page
1.	GRADE LEVELS BY SCHOOL TYPE	8
2.	CLASSROOM TYPES	8
3.	PERCENTAGE OF STUDENTS ON FREE LUNCH	9
4.	TEACHER PERCEPTIONS OF CLASS CONDITIONS	10
5.	TOTAL NUMBER OF STUDENTS IN EACH CLASS	11
6.	NUMBER OF MALE STUDENTS	12
7.	NUMBER OF FEMALE STUDENTS	13
8.	NUMBER OF STUDENTS IN UPPER 25% ACADEMICALLY	14
9.	NUMBER OF STUDENTS IN MID 50% ACADEMICALLY	15
10.	NUMBER OF STUDENTS IN LOWER 25% ACADEMICALLY	16
11.	NUMBER OF STUDENTS IN UPPER 25% OF SES	17
12.	NUMBER OF STUDENTS IN MID 50% OF SES	18
13.	NUMBER OF STUDENTS IN LOWER 25% OF SES	19
14.	NUMBER OF GIFTED STUDENTS	20
15.	NUMBER OF LEARNING DISABLED STUDENTS IN CLASSROOM	21
16.	NUMBER OF ENGLISH SPEAKERS	22
17.	NUMBER OF NON-STANDARD ENGLISH SPEAKERS	23
18.	NUMBER OF SPEAKERS OF ENGLISH AS A SECOND LANGUAGE	24
19.	NUMBER OF STUDENTS WHO SPEAK ONLY A FOREIGN LANGUAGE	25
20.	NUMBER OF YEARS TEACHING	26

21. INSTRUCTIONAL METHOD REVISED TO ONLY THREE CATEGORIES 26

22. Two Observers Observing the Same Ten Teachers . . . 30

23. Total Positive Scores Across Subjects by Norm Groups 35

24. Total Positive Scores Across Arts Subjects by Norm Groups 35

25. Total Positive Scores Across Classroom Types by Norm Groups 36

26. Total Positive Scores Across Ethnic Groups by Norm Groups 36

27. Total Positive Scores Across Teacher's Degree by Norm Groups 37

28. Total Positive Scores Across Subjects by Norm Groups 37

29. Total Positive Scores by SES Characteristics by Norm Groups 38

30. Total Positive Scores Across Sex by Norm Groups . . 38

31. Total Negative Scores Across Subjects by Norm Groups 39

32. Total Negative Scores Across Arts Subjects by Norm Groups 39

33. Total Negative Scores Across Classroom Types by Norm Groups 40

34. Total Negative Scores Across Ethnic Groups by Norm Groups 40

35. Total Negative Scores Across Highest Degree by Norm Groups 41

36. Total Negative Scores Across Experience by Norm Groups 41

37. Total Negative Scores by SES Characteristics by Norm Groups 42

38. Total Negative Scores Across Sex by Norm Groups . . 42

39. UNADJUSTED TOTAL SCORES ACROSS INSTRUCTIONAL METHOD 44

Appendix A

Instruments and Resultant Variables

The FPMS Summative Observation Instrument

SUMMATIVE OBSERVATION INSTRUMENT

Florida Performance Measurement System
Coalition for the Development of a Performance Evaluation System
Office of Teacher Education, Certification and Inservice Staff Development
Tallahassee, Florida

OBSERVATION INFORMATION

Teacher's Name _____ Observer's Name _____

Date _____ Observation Number _____

School _____ Dept./Team _____

Starting Time _____ Length of Observation _____

The summative instrument is designed to record effective and ineffective observable indicators of teacher behavior in four of six research-based domains of effective teaching identified by the Florida Coalition. The indicators included in this instrument are representative of Domains 2.0 Management of Student Conduct, 3.0 Instructional Organization and Development, 4.0 Presentation of Subject Matter, and 5.0 Communication: Verbal and Nonverbal.

DIRECTIONS:

1. Place a mark in the appropriate box when a relevant behavior is observed. Effective items are on the left of the instrument and ineffective items are on the right.
2. Mark items each time they are observed.
3. Sum frequencies by indicators for each domain and record an effective and ineffective total in the Observation Summary.

INSTRUCTION FORMAT

		Planned	Observed	Recommended Changes
Lesson Type:	Introductory			
	Review			
	Developmental			
	Practice			
	Other			
Grouping(s):	Total Group			
	Subgroups			
	Individual			
	Other			
Method(s):	Lecture			
	Discussion			
	Recitation			
	Interaction			
	Independent			
	Disparate			
	Other			

OBSERVATION SUMMARY

Domain	Effective	Ineffective
2.0 Management of Student Conduct		
3.0 Instructional Organization and Development		
4.0 Presentation of Subject Matter		
5.0 Communication: Verbal and Nonverbal		

SUMMATIVE OBSERVATION INSTRUMENT

DOMAIN	EFFECTIVE INDICATORS		FREQUENCY	FREQUENCY	INEFFECTIVE INDICATORS
3.0 INSTRUCTIONAL ORGANIZATION AND DEVELOPMENT	1. Begins instruction promptly				Delays
	2. Handles materials in an orderly manner				Does not organize or handle materials systematically
	3. Orients students to classwork/maintains academic focus				Allows talk/activity unrelated to subject
	4. Conducts beginning /ending review				
	5. Questions: academic comprehension/ lesson development	asks single factual (Dom. 5.0)			Poses multiple questions asked as one, unison response
		requires analysis/reasons			Poses non-academic questions/non-academic procedural questions
	6. Recognizes response/amplifies/gives corrective feedback				Ignores student or response/expresses sarcasm, disgust, harshness
	7. Gives specific academic praise				Uses general, non-specific praise
	8. Provides for practice				Extends discourse, changes topic with no practice
	9. Gives directions/assigns/checks comprehension of homework, seatwork assignment/gives feedback				Gives inadequate directions/no homework/no feedback
	10. Circulates and assists students				Remains at desk/circulates inadequately
4.0 PRESENTATION OF SUBJECT MATTER	11. Treats concept — definition/attributes/examples/non-examples				Gives definition or examples only
	12. Discusses cause-effect/uses linking words/applies law or principle				Discusses either cause or effect only/uses no linking word(s)
	13. States and applies academic rule				Does not state or does not apply academic rule
	14. Develops criteria and evidence for value judgment				States value judgment with no criteria or evidence
5.0 COMMUNICATION: VERBAL AND NONVERBAL	15. Emphasizes important points				
	16. Expresses enthusiasm verbally/challenges students				
	17.				Uses vague/scrambled discourse
	18.				Uses loud-grating, high pitched, monotone, inaudible talk
	19. Uses body behavior that shows interest — smiles, gestures				Frowns, deadpan or lethargic
2.0 MGT. OF STD. CONDUCT	20. Stops misconduct				Delays desist/doesn't stop misconduct/desists punitively
	21. Maintains instructional momentum				Loses momentum — fragments non-academic directions, overdwells

Appendices Section II, A-D

Development of New Domains of Performance

Section II, Appendix A

DOMAIN

6.0 TESTING: STUDENT PREPARATION, ADMINISTRATION, FEEDBACK

6.0 TESTING: STUDENT PREPARATION, ADMINISTRATION, FEEDBACK

Definition: This domain = development and maintenance of an environment in which students can validly demonstrate their knowledge, skills, etc., and receive adequate information about the quality of their test performance.

This domain includes the following named concepts:

6.1 Preparation for Testing

6.2 Test Administration

6.3 Formative Feedback

TESTING

Overview

Testing is an integral part of instruction. Test data are essential for the teacher in making day to day decisions and for the learner in understanding his capacities, achievement, and learning difficulties.

Because tests can have a decided influence upon students, they should not be taken lightly or given haphazardly. To assess student progress, the teacher must have knowledge of how to plan and construction tests, treatment of which is found in the Handbook on Measurement and Evaluation. However, this is not sufficient. The teacher must also know how to administer tests, how to prepare students in the knowledge and skills of test-taking, and how to provide feedback on test performance. These conditions of successful testing are treated in this Domain.

Research studies indicate that students do their best when they are motivated and when they are aware of, and apply, test-taking skills. It is therefore important that teachers motivate students to put forth their best efforts and that they prepare them in the skills of test-taking. Furthermore, cheating has become a major problem in today's schools. While not all cheating can be attributed to inadequate physical setting and poor classroom supervision, these factors can be important in the reduction of cheating.

The use that teachers make of test data is a significant aspect of instruction. Teachers who keep the results of test to themselves reduce the motivation of their students and deny to them information about their own successes and failures and what they need to do to improve their achievement. Frequent testing and feedback to the learner can be, and typically is, an effective way to increase achievement. Frequent feedback is particularly beneficial to learners from poor educational backgrounds and to students with little motivation.

This Domain deals with the foregoing components of testing. It sets forth research knowledge, the application of which can result in reduced student fear of examinations, improved test motivation and test-taking skills, reduced cheating, and a feeling by students that they have been fairly judged.

6.1 Concept: Preparation for Testing

Definition: Preparation for testing = teacher verbal behavior and actions prior to testing that ready the students for tests.

Indicators

6.1.1 Orients Students to Test

Definition: Orients students to test = teacher's verbal behavior that informs students of the purpose of the test, how the test results will be used, and how the results are relevant to them personally.

Example (Pos): Mr. Chen explains to his typing class that every day they will have timed tests. The test scores are to be recorded on individual charts to help them improve their typing rate. The daily scores will not be part of their class grade, but he will look at the scores to guide their improvement. He assures them they can all increase their typing rate with practice.

Example (Pos): Mr. Benson is administering the Stanford Diagnostic Mathematics Test to his six grade class. Mr. Benson states: This test is designed to help me tell which math concepts we have been studying that you do not understand. I will use the results to place you in groups so that I can better help you to master those concepts which are not yet clear to you. The results will not be used in determining your grade, but rather to help us be sure that when you go on to 7th grade math you do know 6th grade math!

Example (Neg): Ms. Jensen tells her class she is going to give them an achievement test tomorrow to find out who should be making the "A's" and "B's" and who should get "F's." Furthermore, if anyone doesn't "measure up" on the test she might confer with that student's parents.

6.1.2 Specifies Test Content

Definition: Specifies test content = teacher tells the class what knowledge is important for them to study and suggests references or guides.

Example (Pos): A few days before the home economics test, Ms. Hernandez reviews the food groups covered in class. She gives the students a study guide containing study questions to take home. She specifies what chapters and topics will be covered on the examination, and tells the class that the problems on the test will be similar to those they have been working on.

Example (Neg): Ms. Wilson tells the students the test they are to take tomorrow includes all kinds of questions about everything they should have learned, so they had better take their books home tonight and study hard.

6.1.3 Instructs Students in Test-Taking

Definition: Instructs students in test-taking = teacher behavior which increases the students' abilities to utilize the characteristics and format of a test and/or test-taking situation to improve their chances of making a higher score.

Example (Pos): In an English class, Mr. Graff instructs his students in how to take a test. He tells them to read the directions and begin work immediately; to mark the items they know, omitting the hard items; to then go back to the hard ones and by progressive elimination, choose the best answer; to check all answers when test is finished. He explains how to use deductive reasoning to choose between options and provides practice in this process.

Example (Neg): Mr. Moto tells his class that there is no way to answer the test questions without mastery of the subject. Study hard. If you don't know the answer, then do the best you can.

Principles

If the teacher informs students of the purpose of the test, how the results will be used and why the results are relevant to them personally, then their test performance will likely improve.

If the teacher instructs students in utilizing the characteristics and formats of tests and/or test-taking situations to receive higher scores, then the tests will be more valid measures of what they are intended to measure.

Supportive Evidence

The perceived importance of a test appears to affect student performance on the test. Anxiety can be a related factor. See Supportive Evidence under 6.2, Test Administration.

Yamamoto and Dizney (1965) found that elementary and secondary students, particularly girls, did better on the average when a test was called an "intelligence" test than they did when the test was called an "achievement" or "routine" test. Naughton (1968) informed a third of his 10th grade class as to how their scores on the Differential Aptitudes Test were to be used and stressed the importance of maximum effort on the test.

The fact that teacher behavior before a test can influence student performance was demonstrated in a study by Gorden and Durea (1948). Scores for a group of students who were discouraged by the examiner were significantly depressed, one-third of a standard deviation on the average, compared to scores of a group that was not discouraged.

The effect of expectancy is supported in a study by Osler (1954). She divided 15 high school classrooms into three groups. One group was informed that they were among the lowest 10% on a pretest in arithmetic, another group that they were the most successful, and a third group given no information. When a test was administered to all groups, she found that students in the success and control groups did significantly better than students in the failure group. Apparently when students are led to perceive themselves as failures, they function that way.

Teachers can help students get ready for a test by giving them specific information about the content to be covered and the type of test to be given. High school students given training in notetaking which emphasized distinguishing between subordinate and superordinate information, abbreviating words, paraphrasing and using an outline format, and who expected a multiple choice test, did significantly better than those who did not receive training; furthermore, they took more efficient notes. In the group who expected an essay test, subjects with no training scored higher than trained students, but among those trained, the students who expected a multiple choice test were better prepared and scored higher on the essay test than those who expected an essay test (Carrier and Titus, 1981).

Test anxiety was reduced and achievement was higher for a group of high anxiety college students who were given study guides for exam preparation. This was true for high, medium and low ability levels according to Szafran (1981).

Another way teachers can help students prepare for tests is to teach them how to take a test. Instruction might include format, content clues, use of test time strategy, etc., and could include practice particularly if separate answer sheets are to be used. Test-taking skills are widely recognized as a source of variance in test scores and as a possible depressor of test validity. Testwiseness permits some students to score higher on classroom and standardized tests than other students who lack test-taking skills, although the students may be equally knowledgable of the subject matter.

Oakland (1972) identified prerequisite abilities needed for taking the Metropolitan Readiness Test. Children from Head Start classes were randomly divided into experimental and control groups. The experimental group was given instruction on these skills (not on the content of the test) during 30-minute periods twice a week for six weeks. The experimental group scored significantly higher on the MPT immediately following the instruction but lost the advantage over the summer months for a second post-test given in September.

Kalechstein, Kalechstein and Docter (1981) report the effects of instruction in test-taking techniques including how to follow oral directions, how to mark responses, strategies for using test time, and guessing strategies, upon standardized reading test scores of second grade Black students. The experimental group which received the instruction performed significantly higher ($p < .01$) than the control group when administered the Stanford Achievement Test.

To determine the effect of test-taking skill training on performance on a school readiness test, an experimental group of Mexican-American kindergarten students received two twenty minute sessions of training. Dreisbach and Keogh (1982) found significantly higher scores for those students who received training, especially boys, regardless of whether the test was administered in Spanish or in English.

Slakter, Koehler, and Hampton (1970) showed that test-taking skills increased as children moved through the grades.

Extensions/Exceptions

One study indicates that emotional factors may play a part in test scores. Certainly the death of a loved one, the anticipation of a birthday celebration, or school experiences may be a positive or negative influence on test performance. Examinees' experiences immediately preceding a test affected their scores in studies by McCarthy (1944) and Reichenberg-Hackett (1953). In the McCarthy study, third and fourth grade children who wrote a composition on "The Best Thing That Ever Happened to Me" prior to a test averaged four to five points higher than their scores on the same test after writing on "The Worst Thing That Ever Happened to Me." Children in the Reichenberg-Hackett study showed significant improvement in their scores after a gratifying experience involving the successful solution of an interesting puzzle, followed by a reward of toys and candy. Improvement did not follow neutral or less gratifying experiences. Other benefits from the positive teacher behavior cited here include better attitude toward schoolwork, not just testing. Shannon (1980) found that student attitudes toward reading were more positive after taking a standardized reading achievement test when the students had been counseled on the purpose and nature of the testing before the test was administered.

It appears that teachers should be positive and approving in a test situation and convey the idea that there is an important task to be accomplished. Pupils should not be threatened with a test or warned of dire consequences if they fail.

Kirkland (1971) reviewed seven studies on the relation of expectation of success or failure to successful performance on tests. Expectation of success or failure on tests determined the amount of time students persisted in attempting to solve problems. Those who had high expectancy of success performed better than those who had a low expectancy.

Typically, teachers encourage students to skip over difficult items and return to them if time permits. Rindler (1980) reported on the correlation of skipping difficult items on a verbal aptitude test for low, middle and high ability students under varying time limits. For the high ability students, subjects who skipped items scored higher than those who did not. However, in the middle and low ability groups, students who skipped items performed more poorly than those who did not.

6.2 Concept: Test Administration

Definition: Test administation = teacher performance that provides favorable environmental conditions for test taking.

Indicators

6.2.1 Arranges Physical Setting

Definition: Arranges physical setting = teacher performance that sets up the physical arrangements (seating, and control of distractions and conditions of temperature, ventilation and lighting) favorable to student performance on examinations.

Example (Pos): The day of the announced examination for his class, Mr. Bloom entered his classroom early. Before the students entered he arranged the chairs to provide as much space as possible between students. He checked the temperature and ventilation and turned on all of the lights in the room. He placed a sign on the door which read "Testing in Progress."

Example (Neg): Ms. Vaughn enters the classroom and says: Okay, class. As you know you're going to have a test today on the last three chapters. We're going to have to move the desks. Please move your desk out of the circular formation and set them up in rows again. John, run down to the custodian's office and ask him if he can come up and do something with that flickering light this morning.

6.2.2 Make Positive Statements

Definition: Makes positive statements = teacher verbal expressions that set a warm and encouraging classroom atmosphere for the examination.
Example (Neg): Ms. Jacobs is faced with the task of administering the State Assessment tests to her 5th grade class. She stands before her class and states: Now you are going to be sorry you haven't been listening to me all year! If you don't pass this test, you won't go on to 6th grade next year. I bet I'll have half of you in my class again. Heaven forbid! Well, do the best you can. You'll have to work fast or you won't get through.

Example (Pos): Ms. Hudson's fifth grade class is scheduled for the State Assessment tests. Ms. Hudson greets her students cheerfully and states: You know that we have been working hard on what is in this test. I am sure you are well prepared and will do well on the test. Do your best and we will all be proud of the results!

6.2.3 Monitors Examination

Definition: Monitors examination = teacher behavior that tells students that he/she knows what they are doing.

Example (Neg): Mr. Woods is proud of his militarily run classroom, his knowledge of the subject matter, his difficult assignments, and his letting students know "their place." When the test booklets have been distributed, Julie raises her hand. MR. WOOD's responds: Julie, it is too late for questions now. After testing has begun, Ms. Allison from next door calls Mr. Wood from the room. They converse in the hall while their classes work on their tests.

Example (Pos): Mr. Page has reviewed the material for his first test and has answered questions to clarify students' understanding. He has invited students who do not feel ready for the test to see him after school for more help. When the students come into the classroom they find the chairs spaced as far apart as possible. Different forms of the test (on two colors of paper) are distributed. Mr. Page places a "Testing in Progress" sign on the door. He projects a friendly, pleasant manner, but at the same time communicates to students that he is alert to what they are doing.

Principles

If tests are administered in a physical setting which is comfortable and free of distractions and opportunities to cheat, by an examiner who is positive and encouraging, then students will have a fair chance to demonstrate what they know and the teacher will more likely obtain a valid measure of their achievement.

If the teacher is aware of the anxiety caused by tests in some students and refrains from using tests as a threat or from emphasizing the negative consequences of poor performance, then high anxious students will suffer less anxiety and perform better on tests.

Supportive Evidence

Walker, Sannito and Firetto (1970) reported that tension oriented instructions, such as telling examinees that everyone is expected to finish in the time allotted, had a beneficial effect on the performance of low anxious persons, but a deleterious effect on that of high anxious persons.

That debilitating anxiety can be induced by the examiner has been demonstrated in several studies. Wrightsman (1962) administered a timed measure of intelligence to two groups of college freshman under two conditions: 122 subjects were led to believe that performance on the test was very important and that the results would follow them throughout their college careers; another 112 subjects were led to believe the test was given only to collect norms. When the test was seen as important, anxious subjects scored significantly lower than non-anxious subjects. In the other condition, anxiety was unrelated to test performance.

For a test situation it is expected that students have comfortable seats of an appropriate size and adequate working space in a well-lighted, well-ventilated room with a minimum of outside noise. However, research on physical setting indicates that some latitude in these conditions can be tolerated without grave consequences. Traxler and Hilkert (1942) reported that high school students who used desks tended to obtain higher test scores than students using chairs with desk arms. However, Ingle and De Amico (1969) found that high school students who took tests with lap boards as a writing surface did no worse on a standarized achievement test than did students who had adequate writing surfaces.

Super, Braasch and Shay (1947) explored the effects of a variety of distractions on test performance of college students. The Otis Test of Mental Ability and two other tests were administered to thirty students in the distracted group and to twenty-six in the control group. The distracted group was subjected to a trumpeter playing in the next room, the examiner's friend bursting into the room and whispering loudly, "How long are you going to be here?", a loud argument outside the classroom door, and mistiming of the test with instructions for the test takers to proceed after the time clock bell rang. Comparison of the scores of the two groups of students found no statistically significant differences. The conclusion was drawn that commonly occurring distractions do not affect test results.

However, distractions may have an impact on scores of younger students. Trentham (1975) studied 72 sixth grade students randomly selected from three elementary schools. The students were randomly divided into experimental and control groups. The distractions used with the experimental group were an alarm bell, two kittens, blinking of the lights, a person entering the room to deliver a note, two talking teachers entering the room, and a radio playing the World Series for 30 seconds. The control group took the test undisturbed. The disturbed group scored significantly lower than the undisturbed control group. This finding supports the desirability of non-distracting testing conditions, at least for younger students.

Bushway and Nash (1977) reviewed the research on school cheating. They cite the studies of McQueen (1957) and Rogosin (1951) which conclude that situational factors in school are more important than behavioral characteristics of students as determinants of whether or not cheating occurs. These situational factors include the teacher's leaving the room during testing, the emotional tone of the classroom, and the moral climate of the school.

Fischer (1970) reported that 65% of the 135 boys and girls in a study in fourth, fifth, and sixth grades cheated under control conditions where students were given only usual test-taking instructions. In the experimental group 76% cheated under the conditions of appeal to honesty where the teacher explained the need for true measures of student's knowledge and effectiveness of her teaching techniques. Thirty-five percent cheated following a discussion of what students think about cheating and signing their names and writing reasons why they would not cheat on a test.

Vitro and Schoer (1972) found that the highest incidence of cheating occurred among students who were unlikely to do well on the test, who were unlikely to get caught, and to whom the particular test was very important.

Uhlig and Howes (1967) found that a large percentage of college students would cheat even in a non-pressure situation were the opportunity present.

Personality and teaching style of the teacher were found to be other situational determiners in cheating. Shirk and Hoffman (1961) found that teachers who are very authoritarian, know all the answers and infer that their students are inferior beings lead student to cheat more. Weldon (1966) found a positive relationship between amount of cheating and totalitarian procedures of teachers. Johnson and Klores (1968) and Steininger, Johnson and Kirts (1964) found that a dissatisfying classroom situation was judged by students as producing a greater amount of cheating.

Extensions/Exceptions

Several studies of test anxiety report 25-30% of the children studied as highly test anxious (Cotler and Palmer, 1970; Nottelman and Hill, 1977). Wine (1971) and Nottelman and Hill have suggested that the debilitating effects of test anxiety may have an attentional explanation. They report that high anxious students seem less able to attend to task-relevant cues, frequently look away from the task and perform more poorly. Wine suggested that high anxious students become preoccupied with (task-irrelevant) worry responses and do not devote enough attention to the test-taking task.

Culler and Holahan (1980) investigated the relation of test anxiety to academic performance in college freshman and the differences in study-related behaviors between high and low anxious individuals. They reported a significant decrement in grade point average associated with test anxiety (mean GPA for low-test anxious group was 2.86; for the high test anxious group, 2.51). High test anxious students were found to have poorer study skills.

Trentham (1979) did not find anxiety-producing instructions to sixth graders to be significant, but the high anxiety students did score significantly lower than medium and low anxiety students.

A comparison of SAT test scores of high school seniors by French (1962) did not indicate a significant effect of anxiety on validity. The only significant effect was high anxiety girls tended to do better on the mathematics test.

Green (1981) reported higher anxious students in statistics made more response changes. Correlations among anxiety, study skills and achievement test scores of college students reported by Kirkland and Hollandsworth (1979) show anxiety is negatively related to grade point average ($r = -.37$) and ACT scores ($r = -.31$). Effective Study Test scores and grade point average positively correlate (.37) ACT scores correlate .40 with Effective Study Test scores.

Smith (1969) reports that white second grade girls have significantly higher test anxiety than white second grade boys, but there was no difference in the test anxiety scores of black boys and girls. The differences in test anxiety scores between boys and girls increases through elementary school.

Bauer (1975) found that boys were more inhibited by anxiety producing instructions than girls. Clawson, Firment and Trower (1981) studied the effects of test anxiety on 150 ninth graders and 112 seventh graders in Northeast Florida. The students were administered the Test Attitude Inventory and the Comprehensive Test of Basic Skills. Mean scores for seventh and ninth grade blacks, in both language and math, centered on the 50th percentile. Whites scored at the 65th percentile and above on all measures. Blacks had significantly higher scores than whites in state, trait, and total test anxiety, by separate seventh and ninth grade groups, as well as in the combined group.

High state (transitory, situational) anxiety was found in 32% of the blacks and only 9.8% of the whites. On the other hand 19% of the whites had low trait (relatively stable characteristic tendency toward anxiety) test anxiety, compared to only 5.3% of blacks. Seven percent of all blacks and 14% of the whites were low in state test anxiety. Clawson et al contend that these findings must be considered in standardized testing, especially where results are used as a basis for graduation, promotion, and remediation.

The teacher may choose from alternative methods of presenting test items without significantly affecting test results. This can make it possible to provide test variety and meet the needs of exceptional children too. Bligh (1965) reviewed several studies which have shown that tests administered via television (Burr, 1963; Curtis and Kropp, 1961) or overhead projector (Schwarz, 1967) yield results comparable to those obtainable on conventionally administered tests, provided that speed is not a factor, and where other conditions are held constant. Burr (1963) studied scores sixth

graders earned on the social studies test of the Metropolitan Achievement Tests. He used three modes of item presentation: projected silence, audio-reinforced, and conventional administration. The experimental techniques used less time than the conventional approach, however, test scores were not significantly different for any test format. The ninth graders tested by Curtis and Kropp (1961) scored significantly higher on items involving word problems that were televised, but on quantitative items which gave computational examples, the normal administration produced significantly higher scores. Ninth graders had no difficulty with verbal abilities items that were televised. In fact, a significantly greater number of total items were responded to on the experimental administration.

In a study of in-class tests versus take-home exams, Marsh (1980) found significantly better results in all categories of the in-class tests. These students also indicated that they had studied significantly more. Kalish (1958) reported no significant difference in total errors between open-book and closed-book exams.

The effect of examiner's race in group tests has been studied by Jensen (1974). An elementary school with 60 percent white and 40 percent black students were given five tests. Twelve white and eight black examiners, administered the tests. The results of the five tests used showed the race of the examiner did not produce a significant impact on the scores of either black or white students.

Mishra (1980) found no significant effect of examiners' ethnicity on Anglo third grade students. Mexican-American students, however, scored significantly higher on vocabulary tests when they were administered by a Mexican-American examiner, but there were no significant differences on other tests.

Intelligence tests administered in English versus in Spanish or in English with an interpreter were composed by Swanson and Deblassie (1979). Spanish-speaking Mexican-American first and second grade children scored significantly higher on the Verbal tests when administered in English; however, in general, their scores were higher on the Performance tests with the Spanish administration. The authors discuss various influences which may have contributed to these findings such as the unfamiliarity with the use of Spanish in the scholastic setting leading to reticence on the part of the children. Other possible factors include the lack of direct Spanish counterparts to many English concepts and phrases, and the fact that facility in Spanish was an uncontrolled variable. The superiority of the Spanish administration of the Performance tests was hypothesized as attributable to decreased anxiety in the children.

Results of a non-verbal (perceptual/spatial) intelligence test with instructions given in English versus in Spanish were compared for bilingual Puerto-Rican sixth, seventh and eighth grade students (Anastasi and Cordova, 1953). No significant effects were found for language used in the instructions, order of the language used in the presentation or sex.

6.3 Concept: Formative Feedback

Definition: Formative feedback = teacher behavior that provides information to students about their test performance.

6.3.1. Reviews Tests

Definition: Reviews test = teacher uses test data to identify and correct errors in understanding.

Example (Pos): Ms. Thomas is aware of need for specific feedback concerning errors. The day after the test she scheduled a discussion session to go over the items on the test. She listened carefully to reasons students gave for alternative responses to questions. She identified areas of misunderstanding and clarified misconceptions.

Example (Neg): Mr. Zayer gives a test to his history class, saying he will return the tests the next day. The tests were actually returned a week later with grades and no comments. Students ask about the answers to certain questions. He tells them to refer to their textbook.

6.3.2 Comments Positively on Test Results

Definition: Comments positively on test results: teacher's verbal behavior that expresses approval of student response to test item(s) or to discussion of responses.

Example (Pos): Mr. Crawford returns test papers and calls attention to items most frequently missed. He asks some students to explain their answers and during the class discussion Mr. Crawford recognized different points of view and commented on the students' good thinking.

Example (Neg): Ms. Jenkins, on returning test papers, answers student explanation of their answers with such remarks as "How could you think that," "You must not have studied," "We discussed that thoroughly in class; you must not have listened."

Principle

If the teacher provides feedback to students on their tests, then motivation, learning, and retention will increase, and the teacher will be more able to adjust instruction to the needs of the classroom group.

Supportive Evidence

Students receive constant feedback from teachers, classmates and others in the school. The concern in this review is only with the effects of feedback from formative testing on later school achievement, retention and attitudes toward the subject matter.

Barringer and Gholson (1979) reviewed more than seventy studies on the effects of feedback on concept formation and concept identification in children. Three combinations of feedback were investigated: (a) feedback which follows both correct and incorrect responses (right-wrong); (b) feedback which follows correct responses, but does not follow incorrect responses (right-blank); and (c) feedback which follows incorrect responses, but does not follow correct responses (wrong-blank). They reported: that "In general, the research has been fairly consistent in two respects: (a) verbal and symbolic feedback produce more rapid acquisition than does tangible feedback, and (b) the right-blank combination is least efficient as a method for teaching children conceptual material. Wrong-blank has usually resulted in the most efficient performance during acquisition and the greatest resistance to extinction when all feedback was discontinued. The relative efficacy of right-wrong has been less consistent. In some cases, this combination has been similar, in effect, to wrong-blank; in others, it has been intermediate or similar to right-blank."

Reinforcing only correct responses is ineffective because children interpret the nonoccurrence of feedback as positive feedback. When no feedback follows incorrect responses they repeat their mistakes (Mims and Gholson, 1977).

Token economy has been effectively used to modify a variety of classroom social behaviors (aggression, social withdrawal, and inappropriate talking). Tangible feedback is less facilitative of conceptual learning, however, than is verbal or symbolic feedback. Some investigators (Mims and Gholson, 1977; Ratliff, 1972; Spence, 1970) have suggested that tangible feedback distracts the child's attention from relevant stimulus cues.

Hughes (1973) studied the effects of the quality of feedback on seventh grade students in 13 classrooms. The teacher planned the lesson in detail and memorized the presentation. Lesson content was held constant while teacher reactions to student responses were manipulated. The teacher either praised or reacted neutrally. Subsequent testing showed that positive teacher reactions facilitated student achievement significantly more than minimal (neutral) reactions.

Page (1958) demonstrated the importance of specificity of feedback. In 74 classes teachers administered an objective test to their students. A score and a grade was assigned to each paper and each paper was randomly assigned to one of three groups. Group I papers received no comment other than a score and a grade; Group II papers received general and encouraging comments, like "Good work, keep it up;" and Group III received specific comments that the teacher thought would be helpful. The subsequent criteria test, an objective test, showed that the highest scores were achieved by pupils who received specific comments, the next highest by those who received general comments, and the lowest scores by those who received no comments.

In an investigation with 233 seventh-grade students, Bridgeman (1974) found that "success" feedback was followed by significantly higher

performance than was "failure" feedback in students who had performed equally well on an earlier test. Scores from a test administered two days earlier were randomly assigned to the students immediately before taking the test. The results suggest that the student's perception of his prior success or failure influences his subsequent performance.

It is more effective to delay feedback from a test if one wants the greatest retention of learning. Brackbill and Kappy (1962) found that immediate reinforcement, as received on programmed materials or on a computer test, is not as effective as delayed feedback. After a test, it is better to wait at least 24 hours to go over the items. When feedback from a test is delayed by several days, learners perform better on a retest than when feedback is provided immediately or after 24 hours. Presumably, wrong answers interfere with learning and recalling the correct answer provided by the immediate feedback (Surber and Anderson, 1975).

Extensions and Exceptions

In reinforcement studies of lower class black males aged 9-11 years, Terrill, Terrill and Taylor (1980, 1981) found tangible reinforcement was more effective than other types of reinforcement and no reward. With mildly mentally retarded black males both the candy and culturally relevant praise such as "nice job, little brother" significantly increase their test scores.

Institutionalized educable mentally retarded students who received verbal praise and reinforcement by positive examiners scored significantly higher on the WISC-R Performance (Saigh, 1981).

Eighth grade social studies students who received knowledge of the correct response within a short interval after the response was emitted were compared with subjects who did not have knowledge of the correct response. Oliver and Taylor (1970) found higher retention scores for slow students who received feedback, but no significant difference for accelerated students.

REFERENCES

Anastasi, A., and Cordova, F.A. Some Effects of Bilingualism upon the Intelligence Test Performance of Puerto Rican Children in New York City. Journal of Educational Psychology (1953) 44:1-19.

Barringer, C., and Gholson, B. Effects of Type and Combination of Feedback upon Conceptual Learning by Children: Implications for Research in Academic Learning. Review of Educational Research (1979) 49:459-478.

Bauer, D. The Effects of Instructions, Anxiety and Locus of Control on Intelligence Test Scores. Measurement and Evaluation in Guidance (1975) 8:12-19.

Bligh, H. F. Trends in the Measurement of Educational Achievement. Review of Educational Research (1965) 35:34-49.

Brackbill, Y., and Kappy, M. S. Delay of Reinforcement and Retention. Journal of Comparative and Physiological Psychology (1962) 55:14-18.

Bridgeman, B. Effects of Test Score Feedback on Immediately Subsequent Test Performance. Journal of Educational Psychology (1974) 66:62-66.

Burr, W. L. Empirical Relationships Among Modes of Testing, Modes of Instruction and Reading Levels: In Sixth Grade Social Studies. Journal of Experimental Education (1963) 31:433-435.

Bushway, A., and Nash, W. R. School Cheating Behavior. Review of Educational Research (1977) 147:623-632.

Carrier, C. A., and Titus, A. Effects of Notetaking Pretraining and Text Mode Expectations on Learning from Lectures. American Educational Research Journal (1981) 18:385-397.

Clawson, T. W., Firment, C. K., and Trower, T. L. Test Anxiety: Another Origin for Racial Bias in Standardized Testing. Measurement and Evaluation in Guidance (1981) 13:210-215.

Cotler, S., and Palmer, R. The Effects of Text Anxiety, Sex of Subject, and Type of Verbal Reinforcement on Maze Performance of Elementary School Children. Journal of Personality (1970) 38:216-234.

Culler, R. E., and Holahan, C. J. Text Anxiety and Academic Performance: The Effects of Study-Related Behaviors. Journal of Educational Psychology (1980) 72:16-20.

Curtis, H. A., and Kropp, R. P. A Comparison of Scores Obtained by Administering a Test Normally and Visually. Journal of Experimental Education (1961) 29:249-260.

Dreisbach, M., and Keogh, B. K. Testwiseness as a Factor in Readiness Test Performance of Young Mexican-American Children. Journal of Educational Psychology (1982) 74:224-229.

Fischer, C. T. Levels of Cheating under Conditions of Informative Appeal to Honest, Public Affirmation of Value and Threats of Punishment. Journal of Educational Research (1970) 64:12-16.

French, J. W. Effect of Anxiety on Verbal and Mathematical Examination Scores. Educational and Psychological Measurement (1962) 22:553-564.

Gorden, L. V., and Durea, M. A. The Effect of Discouragement on the Revised Stanford-Binet Scale. Journal of Genetic Psychology (1948) 73:201-207.

Green, K. Item-Response Changes on Multiple-Choice Tests as a Function of Test Anxiety. Journal of Experimental Education (1981) 49:225-228.

Hughes, D. C. An Experimental Investigation of the Effects of Pupil Responding and Teacher Reacting on Pupil Achievement. American Educational Research Journal (1973) 10:21-37.

Ingle, B., and DeAmico, G. Effect of Physical Conditions of the Test Room on Standardized Achievement Test Scores. Journal of Educational Measurement (1969) 6:237-240.

Jensen, A. R. The Effect of Race of Examiner on the Mental Test Scores of White and Black Pupils. Journal of Educational Measurement (1974) 11:1-14.

Johnson, R. E., and Klores, M. S. Attitudes Toward Cheating as a Function of Classroom Dissatisfaction and Peer Norms. Journal of Educational Research (1968) 62:60-64.

Kalechstein, P., Kalechstein, M. and Docter, R. The Effects of Instruction on Test-Taking Skills in Second Grade Black Children. Measurement and Evaluation in Guidance (1981) 13: 198-202.

Kalish, R.A. An Experimental Evaluation of the Open Book Examination. Journal of Educational Psychology (1958) 49: 200-204.

Kirkland, K., and Hollandsworth, J.G., Jr. Test Anxiety, Study Skills, and Academic Performance. Journal of College Student Personnel (1979) 20: 431-436.

Kirkland, M.C. The Effects of Tests on Students and Schools. Review of Educational Research. (1971) 41: 303-351.

Marsh, R. Should We Discontinue Classroom Tests? An Experimental Study. High School Journal (1980) 63: 288-292.

McCarthy, D. A Study of the Reliability of the Goodenough Drawing Test of Intelligence. Journal of Psychology (1944) 18:201-216.

McQueen, R. Examination Deception as a Function of Residual Background and Immediate Stimulus Factors. Journal of Personality (1957) 25:643-650.

Mims, M., and Gholson, B. Effects of Type and Amount of Feedback upon Hypothesis Sampling Among 7-8-Year Old Children. Journal of Experimental Child Psychology (1977) 24:358-371.

Mishra, S. P. The Influence of Examiners' Ethnic Attributes on Intelligence Test Scores. Psychology in the Schools (1980) 17:117-122.

Naughton, J. A Modest Experiment on Test Motivation. Personnel and Guidance Journal (1968) 46:606.

Nottelman, E. D., and Hill, K. T. Test Anxiety and Off-Task Behavior in Evaluative Situations. Child Development (1977) 48:225-231.

Oakland, T. The Effects of Testwiseness Materials on Standardized Test Performance of PreSchool Disadvantaged Children. Journal of School Psychology (1972)10:355-[illegible]0.

Oliver, J. K., and Taylor, B. L. The Effects of Teaching Social Studies on Immediate Retention. Journal of Educational Research (1970) 63:198-200.

Osler, S. F. Intellectual Performance as a Function of Two Types of Psychological Stress. Journal of Experimental Psychology (1954) 47:115-121.

Page, E. B. Teacher Comments and Student Performance: A Seventy-Four Classroom Experiment in School Motivation. Journal of Educational Psychology (1958) 49:173-181.

Ratliff, R. G. Two-Choice Discrimination Learning in Children as a Function of Punishment Modality and Reinforcement Combination. Journal of Experimental Child Psychology (1972) 14:365-371.

Reichenberg-Hackett, W. Changes in Goodenough Drawings After a Gratifying Experience. American Journal of Orthopsychiatry (1953) 23:501-517.

Rindler, S. E. The Effects of Skipping Over More Difficult Items on Time-Limited Tests: Implications for Test Validity. Educational and Psychological Measurement (1980) 40:989-998.

Rogosin, H. What About "Cheating" on Examinations and Honesty? School and Society (1951) 74:402-403.

Saigh, P. A. The Effects of Positive Examiner Verbal Comments on the Total WISC-R Performance of Institutionalized EMR Students. Journal of School Psychology (1981) 19:86-91.

Schwarz, J. C. A New Procedure for Administering Objective Tests to Large Classes. Journal of Educational Measurement (1967) 4:167-168.

Shannon, A. J. Effects of Methods of Standardized Reading Achievement Test Administration on Attitudes toward Reading. Journal of Reading (1980) 23:684-686.

Shirk, E., and Hoffman, R. W. The Academic Setting of the Dishonest Student. Improving College and University Teaching (1961) 9:130-134.

Slakter, M. J., Koehler, R. A., and Hampton, S. H. Grade Level, Sex and Related Aspects of Testwiseness. Journal of Educational Measurement (1970) 7:119-122.

Spence, J. T. The Distraction Effects of Material Reinforcers in the Discrimination Learning of Lower and Middle Class Children. Child Development (1970) 41:103-111.

Steininger, M., Johnson, R. E., and Kirts, D. K. Cheating on College Examinations as a Function of Situationally Aroused Anxiety and Hostility. Journal of Educational Psychology (1964) 55:317-324.

Super, D. E., Braasch, W. F., Jr., and Shay, J. B. The Effects of Distractions on Test Results. Journal of Educational Psychology (1947) 38:373-377.

Surber, J. R., and Anderson, R. C. Delay-Retention Effect in Natural Classroom Settings. Journal of Educational Psychology (1975) 67:170-173.

Swanson, E. N., and Deblassie, R. R. Interpreter and Spanish Administration Effects on the WISC Performance of Mexican-American Children. Journal of School Psychology (1979) 17:231-236.

Szafran, R. F. Question-Pool Study Guides: Effects on Test Anxiety and Learning Retention. Teaching Sociology (1981) 9:31-43.

Terrill, F., Terrill, S.L., and Taylor, J. Effects of Race of Examiner and Type of Reinforcement on the Intelligence Test Performance of Lower-Class Black. Psychology in the Schools (1980) 17: 270-272.

Traxler, A. E., and Hilkert, R. N. Effect of Type of Desk on Results of Machine-Scored Tests. School and Society (1942) 56:277-296.

Trentham, L. L. The Effect of Distractions on Sixth-Grade Students in a Testing Situation. Journal of Educational Measurement (1975) 12:13-18.

Trentham, L. L. Anxiety and Instruction Effects on Sixth-Grade Students in a Testing Situation. Psychology in the Schools (1979) 16:439-443.

Uhlig, G. E., and Howes, B. Attitude Toward Cheating and Opportunistic Behavior. Journal of Educational Research (1967) 60:411-412.

Vitro, F. T., and Schoer, L. A. The Effects of Probability of Test Success, Test Importance, and Risk of Detection on the Incidence of Cheating. Journal of School Psychology (1972) 10:269-277.

Walker, R. E., Sannito, T. C., and Firetto, A. C. The Effect of Subjectively Reported Anxiety on Intelligence Test Performance. Psychology in the Schools (1970) 7:241-243.

Wine, J. Test Anxiety and Direction of Attention. Psychological Bulletin (1971) 76:92-104.

Wrightsman, L. S., Jr. The Effects of Anxiety, Achievement Motivation, and Task Importance upon Performance on an Intelligence Test. Journal of Educational Psychology (1962) 53:150-156.

Yamamoto, K., and Dizney, H. F. Effects of Three Sets of Test Instructions on Scores on an Intelligence Scale. Educational and Psychological Measurement (1965) 25:87-94.

Section II, Appendix B

FORMATIVE OBSERVATION INSTRUMENT

6.0 TESTING: STUDENT PREPARATION, ADMINISTRATION, FEEDBACK

Florida Performance Measurement System
Coalition for the Development of a Performance Evaluation System
Office of Teacher Education, Certification and Inservice Staff Development
Tallahassee, Florida

IDENTIFICATION AND FORMAT INFORMATION

Teacher's Name:_____ Date: _____

School: _____ Department/Team: _____

Observer's Name:_____ Observation Number:_____

Starting Time:_____ Length of Observation:_____

Brief Statement of Lesson Purpose or Content:_____

_____

_____

This instrument is designed to record effective and ineffective indicators of teacher behavior in the domain of Testing: Student Preparation, Administration, Feedback. The instrument is divided into three categories: Preparation For Testing; Test Administration; and Formative Feedback.

Directions:

1. Place a mark in the appropriate box when a relevant behavior is observed. (Effective items are on the left of the instrument and ineffective items are on the right.)
2. Mark an item each time it is observed.
3. Sum frequencies by indicators, record subtotals for each category and record an effective and ineffective total for the categories in the appropriate spaces provided below.

Note: Items on the Observation Instrument are observed at varying times. They are as follows:

6.1 Preparation for Testing. Observed one to three days before testing.

6.2 Test Administration. Observed during the period students take an examination.

6.3 Formative Feedback. Observed two or more days after an examination.

Data Summary

CATEGORY	EFFECTIVE	INEFFECTIVE
6.1 Preparation for Testing		
6.2 Test Administration		
6.3 Formative Feedback		
TOTAL	114	

DOMAIN 6.0 OBSERVATION INSTRUMENT

CATEGORY	EFFECTIVE INDICATOR	FREQUENCY	FREQUENCY	INEFFECTIVE INDICATOR
6.1 PREPARATION FOR TESTING (Observed prior to testing.)	States test purpose(s)			
	Explains use of test results			Relates test purpose to grading
	Explains importance of results			Threatens reprisals for failure
	States positive expectations			Makes negative comments about student ability
	States content to be tested			Uses vague/general terms to describe test content or review
	Reviews test content			
	Cites study references or provides study guide			Emphasizes test difficulty
	Describes item type(s)			
	Gives test-taking directions			Makes indefinite statements about test format/directions
	Explains test-taking strategies			
	Provides test-taking practice			
6.2 TEST ADMINISTRATION (Observed during testing period.)	Arranges chairs to provide space among students			Allows close proximity
	Controls temperature and ventilation			Disregards room temperature extremes
	Provides lighting			Disregards shadows, reflections or other poor lighting
	Clarifies student questions			Refuses to help or respond to test related questions
	Offers assistance to individuals with questions			
	Monitors students while taking exam			Leaves the room during testing
	Controls noise and interruptions			Occupied with other tasks/loses awareness of student behavior
6.3 FORMATIVE FEEDBACK (Observed when test results are returned.)	Reviews test items/gives reason for answers			Uses tokens as feedback
	Listens to alternative responses			
	Clarifies misconceptions			
	Makes positive comments about student ideas			

The FPMS Frame Factor Instrument

FPMS

Directions for Completing the

Frame Factor Instrument

1.0 Class Factors

1.1 The subject being taught during observation...math, language, science, p.e., etc.

1.2 The grade level of most of the students in the class Grade 1, 2, 3/10, 11, etc.

1.3 Self-contained = One classroom separated from other classrooms by partitions.
Other = Open classrooms, laboratories, shops, etc.

1.4 Poor = Not conducive to instruction.
Excellent = special facilities provided to complement instruction.

1.5 Sum the number of students who are regular members of the class and record the number of boys and girls.

2.0 Student Factors

2.1 Compare the achievement of your students with standard achievement scores. Then, rank the number of students in your class according to upper, middle or lower achievement.

2.2 Low = The number of students on full or reduced lunch.
Middle and High = Your best judgement of the students' status.

2.3 Gifted = Students who qualify for gifted programs.
Learning Impaired - Students classified in any program such as EMR, EH, etc.

2.4 Standard English - English spoken without dialect.
Non-standard English = English spoken with dialect.
English as second language = Students whose native language is other than English.
Foreign Language only - Non-English speakers.

3.0 Teacher Factors

3.1

3.2

3.3 Count all years of experience whether they are in the same subject or not, private or public education, etc.

3.4 Bachelors, Masters, Educ. Specialist, Doctorate.

FPMS - FRAME FACTOR INSTRUMENT

1.0 CLASS FACTORS:

1.1 SUBJECT _____

1.2 GRADE LEVEL _____

1.3 CLASSROOM: ____ SELF-CONTAINED
____ OTHER (SPECIFY) _____

1.4 CLASSROOM CONDITIONS (LIGHTING, NOISE, CROWDING, ETC.)
____ POOR ____ ADEQUATE ____ EXCELLENT

1.5 CLASS SIZE: _____ # STUDENTS IN TOTAL CLASS
1.5.1 _____ # BOYS
1.5.2 _____ # GIRLS

2.0 STUDENT FACTORS:

2.1 ACADEMIC STATUS: ____ # UPPER 25% OF STUDENT BODY
____ # MIDDLE 50% OF STUDENT BODY
____ # LOWER 25% OF STUDENT BODY

2.2 SOCIO ECONOMIC STATUS: ____ # HIGH SES
____ # MIDDLE
____ # LOW

2.3 EXCEPTIONALITIES: ____ # GIFTED
____ # LEARNING IMPAIRED (EMR,EH,ETC.)

2.4 LINGUISTIC STATUS: ____ # STANDARD ENGLISH SPEAKERS
____ # NON-STANDARD ENGLISH
____ # ENGLISH AS SECOND LANGUAGE
____ # FOREIGN LANGUAGE ONLY

3.0 TEACHER FACTORS:

3.1 SEX: ____ MALE ____ FEMALE

3.2 ETHNICITY: ____ CAUCASIAN ____ HISPANIC
____ AFRO-AMERICAN ____ OTHER

3.3 TOTAL # OF YEARS TEACHING: _____

3.4 HIGHEST DEGREE COMPLETED: _____

Nature and Content of Instruments

The following data gathering instruments were used in this study:

1. The FPMS Summative Observation Instrument,
2. The FPMS Instruction Format subsection of the Summative Observation Instrument, and
3. The FPMS Frame Factor Instrument.

The FPMS Summative Observation Instrument.

This research based observation instrument is the culmination of an intensive five year development effort. The instrument has successfully undergone thorough and rigorous tests of content validity, concurrent validity and reliability (Micceri, 1983, Coalition, 1983).

The instrument consists of 20 positive (effective) items, and 19 negative (ineffective) items, each corresponding to indicators of teacher behavior referenced in the Domain Documents of the FPMS. The items on the Summative Instrument relate to the four observable domains of teacher behavior (Management of Student Conduct, Presentation of Subject Matter, Oraganization of material, and Communication Verbal and Non-verbal). Ten possible scores result from the use of this instrument, one for each domain (positive and negative), and one each for all positive and all negative behaviors.

A retest of the FPMS summative instrument's reliability indicates that the instrument is highly reliable for the positive scores, and adequately reliable for negative scores, despite their infrequent occurence and small variance, when a quartile based scaling is used, if each teacher is observed by at least two observers. The indicators of the Summative Instrument are outlined in Appendix A.

Instruction Format Instrument.

The Instruction Format section of the FPMS is designed to identify the general types of interaction occuring during an observed lesson between teachers and students. It was believed prior to the conduct of this norming study, that the factors identified by this section of the instrument would be important when interpreting observation results. Three general areas of instruction format are obtained with this instrument:

1. Lesson Type
   a) Introductory or Developmental,
   b) Review or practice, and
   c) Combinations of the above.
2. Student Groupings
   a) Total Group
   b) Small Groups or Individual instruction, and
   c) Combinations of the above.
3. Instruction Method
   a) Lecture
   b) Student Teacher Interaction type
      i) Dicussion
      ii) Interaction
      iii) Recitation
   c) Individual
   d) Combinations:
      i) Lecture with Interaction (the most common),
      ii) Lecture with Independent (usually lab work),
      iii) Interaction with Independent, and
      iv) All three in one lesson.

As can easily be seen, Instructional Method (IM) can impact both the type and visibility of teacher behaviors.

Frame Factor Instrument.

The Frame Factor Instrument used in this study is a reduced and simplified version of the instrument used in the Reliability Study conducted during 1982-1983. The instrument obtains teacher perceptions of several variables in each of the following areas:

1. Subject and Class situation,

2. Student Factors such as Academic Status and SES, and

3. Teacher Factors such as experience and training.

Variables

The variables obtained for analysis fall into several categories:

1. Twenty positive (effective) indicators of the Summative Instrument,

2. Nineteen negative (ineffective) indicators of the Summative Instrument,

3. Socio-demographic and Classification variables:

   a) School and Classroom factors

      i) grade level,

      ii) school percentage of students on free lunch,

      iii) class conditions (lighting, noise, etc.), and

      iv) class size

   b) Student Variables:

      i) number of boys and girls,

      ii) teacher perceptions of : student socio-economic status, and academic status,

      iii) number of learning disabled and gifted students,

      iv) number of non-standard English speakers.

   c) Teacher Variables:

      i) experience,

      ii) education,

      iii) sex, and

      iv) ethnicity.

d) Lesson variables:

i) length of lesson,

ii) time of day (morning or afternoon),

iii) class grouping (total group, small group or individual),

iv) lesson type (intrductory, practice, review, etc.),

v) instruction method (lecture, interaction, independent, combinations).

Most of the variables in this study are multipoint ordinal in nature, including all items on the FPMS, all student variables, most school and classroom variables and teacher experience. The remainder may be treated as classification or categorical type variables (nominal level) as independent variables in analysis, or coded into vectors for entry into multiple regression analysis.

All of these variables are dependent upon the accuracy and objectivity of the respondents (teachers and observers). Several are perceptual in nature, particularly SES, academic status, non-standard English, and class conditions. Much of this information can be based upon student records, however, the relative accruacy of responses depends upon both the observer's instructions, and the concern and objectivity of the teacher. To test the accuracy of teacher responses to SES questions, the percentage of students on free lunch was obtained for each school. This was then compared to teacher perceptions of low SES students in their classes and found to correlate highly, $r = .61$, providing some support for the use of teacher perceptions as Frame Factors.

Appendix B

Socio-Demographic, Classroom and Lesson Variables

The following tables, in addition to tables found in the narrative, present the demographic and lesson format aspects of the normative study data base. Each table indicates the variable of interest, the subgroups included in each variable (eg., grade levels - elementary, middle, secondary), the absolute frequency in each subgroup, the absolute percentage each frequency represents of the total data base, the number of missing cases for that variable, and the relative percentage each frequency represents without considering missing cases. The following types and sub-types of variables are found in the data base:

1. school and classroom variables,
2. student variables
    a) academic status
    b) socio-economic status
    c) special students
    d) language status
3. teacher variables
4. lesson variables
5. observation variables

TABLE 1

GRADE LEVELS BY SCHOOL TYPE

	ABSOLUTE FREQUENCY	ABSOLUTE %	RELATIVE %
ELEMENTARY SCHOOL (0 - 5)	468	38.8	39.4
MIDDLE SCHOOL (6 - 9)	318	26.4	26.7
SECONDARY SCHOOL (10-12)	403	33.4	33.9
MISSING	17	1.4	MISSING
	1206	100.0	100.0

VALID CASES 1189 MISSING CASES 17

TABLE 2

CLASSROOM TYPES

	ABSOLUTE FREQUENCY	ABSOLUTE %	RELATIVE %
SELF CONTAINED	989	80.9	82.1
POD OR OPEN	84	6.9	44.0
PE COURT	34	2.8	17.8
LABS	20	1.6	10.5
RESOURCE ROOM	22	1.8	11.5
DEPARTMENTAL	17	1.4	8.9
MEDIA-CLINIC	7	0.6	3.7
AUDITORIUM	1	0.1	0.5
OTHER	6	0.5	3.1
MISSING	19	1.6	MISSING
	1223	100.0	100.0

VALID CASES 1201 MISSING CASES 19

TABLE 3

PERCENTAGE OF STUDENTS ON FREE LUNCH

% OF STUDENTS ON FREE LUNCH	ABSOLUTE FREQUENCY	ABSOLUTE %	RELATIVE %
2%	107	8.7	9.3
6%	4	0.3	0.3
8%	15	1.2	1.3
13%	14	1.1	1.2
14%	110	9.0	9.6
19%	82	6.7	7.2
20%	4	0.3	0.3
21%	30	2.5	2.6
23%	122	10.0	10.6
31%	54	4.4	4.7
32%	64	5.2	5.6
33%	17	1.4	1.5
35%	56	4.6	4.9
36%	19	1.6	1.7
37%	80	6.5	7.0
38%	26	2.1	2.3
41%	63	5.2	5.5
43%	20	1.6	1.7
44%	60	4.9	5.2
45%	24	2.0	2.1
48%	42	3.4	3.7
61%	30	2.5	2.6
71%	20	1.6	1.7
72%	11	0.9	1.0
74%	19	1.6	1.7
80%	29	2.4	2.5
84%	24	2.0	2.1
MISSING	77	6.3	MISSING
	1223	100.0	100.0

MEAN	32.279	MEDIAN	31.984
SKEWNESS	0.724	RANGE	82.000
VALID CASES	1146	MISSING CASES	77

TABLE 4

TEACHER PERCEPTIONS OF CLASS CONDITIONS

	ABSOLUTE FREQUENCY	ABSOLUTE %	RELATIVE %
POOR	86	7.0	7.4
ADEQUATE	764	62.5	65.3
EXCELLENT	320	26.2	27.4
MISSING	53	4.3	MISSING
TOTAL	1223	100.0	100.0

MEAN	2.200	MEDIAN	2.153
SKEWNESS	0.046	RANGE	2.000
VALID CASES	1170	MISSING CASES	53

TABLE 5

TOTAL NUMBER OF STUDENTS IN EACH CLASS

# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT
2.	10	1	1	21.	45	4	33	40.	3	0	98
3.	5	0	1	22.	69	6	39	41.	2	0	98
4.	6	0	2	23.	68	6	45	42.	4	0	98
5.	15	1	3	24.	66	5	50	43.	2	0	98
6.	18	1	4	25.	74	6	56	44.	2	0	99
7.	20	2	6	26.	60	5	61	45.	3	0	99
8.	15	1	7	27.	71	6	67	47.	1	0	99
9.	10	1	8	28.	57	5	72	48.	1	0	99
10.	15	1	9	29.	55	5	76	49.	1	0	99
11.	18	1	11	30.	76	6	83	50.	2	0	99
12.	19	2	13	31.	49	4	87	51.	1	0	99
13.	15	1	14	32.	38	3	90	52.	1	0	99
14.	23	2	16	33.	31	3	93	53.	1	0	99
15.	21	2	17	34.	24	2	95	57.	1	0	100
16.	20	2	19	35.	7	1	95	62.	1	0	100
17.	21	2	21	36.	13	1	96	65.	1	0	100
18.	32	3	23	37.	7	1	97	69.	1	0	100
19.	33	3	26	38.	6	0	97	71.	1	0	100
20.	43	4	30	39.	2	0	97	99.	2	0	100

MEAN	23.677	MEDIAN	24.455
SKEWNESS	0.948	RANGE	97.000
VALID CASES	1208	MISSING CASES	15

TABLE 6

NUMBER OF MALE STUDENTS

# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT
0.	25	2	2	14.	111	9	70	28.	2	0	99
1.	16	1	3	15.	90	8	78	29.	1	0	99
2.	22	2	5	16.	74	6	84	30.	2	0	99
3.	31	3	8	17.	54	5	88	31.	1	0	99
4.	24	2	10	18.	26	2	91	32.	1	0	99
5.	34	3	13	19.	29	2	93	35.	1	0	99
6.	31	3	15	20.	24	2	95	36.	2	0	99
7.	38	3	19	21.	13	1	96	39.	1	0	100
8.	52	4	23	22.	10	1	97	42.	1	0	100
9.	64	5	28	23.	8	1	98	48.	1	0	100
10.	84	7	36	24.	2	0	98	61.	1	0	100
11.	103	9	44	25.	3	0	98	64.	1	0	100
12.	98	8	52	26.	5	0	99	78.	1	0	100
13.	97	8	61	27.	1	0	99				

MEAN	12.137	MEDIAN	12.199
SKEWNESS	2.238	RANGE	77.999
VALID CASES	1185	MISSING CASES	38

TABLE 7

NUMBER OF FEMALE STUDENTS

# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT
0.	38	3	3	14.	91	8	71	28.	3	0	99
1.	30	3	6	15.	77	6	78	29.	2	0	99
2.	40	3	9	16.	68	6	83	31.	1	0	99
3.	36	3	12	17.	44	4	87	32.	1	0	99
4.	35	3	15	18.	45	4	91	34.	1	0	99
5.	23	2	17	19.	30	3	93	38.	3	0	99
6.	40	3	20	20.	24	2	95	40.	1	0	100
7.	35	3	23	21.	10	1	96	41.	1	0	100
8.	54	5	28	22.	9	1	97	42.	1	0	100
9.	72	6	34	23.	6	1	97	45.	1	0	100
10.	94	8	42	24.	4	0	98	64.	1	0	100
11.	100	8	50	25.	4	0	98	68.	1	0	100
12.	81	7	57	26.	2	0	98				
13.	73	6	63	27.	3	0	99				

MEAN	11.521	MEDIAN	11.455
SKEWNESS	1.477	RANGE	67.999
VALID CASES	1185	MISSING CASES	38

TABLE 8

NUMBER OF STUDENTS IN UPPER 25% ACADEMICALLY

# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT
0.	527	49	49	13.	10	1	90	26.	2	0	97
1.	27	3	52	14.	15	1	92	27.	3	0	97
2.	41	4	56	15.	9	1	92	28.	3	0	98
3.	39	4	59	16.	6	1	93	29.	2	0	98
4.	54	5	64	17.	6	1	94	30.	6	1	98
5.	67	6	71	18.	4	0	94	31.	3	0	99
6.	42	4	74	19.	3	0	94	32.	5	0	99
7.	42	4	78	20.	8	1	95	33.	4	0	99
8.	37	3	82	21.	2	0	95	34.	2	0	100
9.	13	1	83	22.	3	0	95	35.	1	0	100
10.	38	4	87	23.	5	0	96	36.	1	0	100
11.	14	1	88	24.	5	0	96	37.	1	0	100
12.	14	1	89	25.	5	0	97	99.	1	0	100

MEAN	4.657	MEDIAN	0.797
SKEWNESS	3.463	RANGE	98.999
VALID CASES	1070	MISSING CASES	153

TABLE 9

NUMBER OF STUDENTS IN MID 50% ACADEMICALLY

# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT.	FREQ	ADJ PCT	CUM PCT	# STUDT.	FREQ	ADJ PCT	CUM PCT
0.	268	25	25	16.	29	3	64	32.	9	1	97
1.	11	1	26	17.	35	3	68	33.	7	1	98
2.	10	1	27	18.	34	3	71	34.	4	0	98
3.	7	1	28	19.	21	2	73	35.	1	0	98
4.	14	1	29	20.	36	3	76	36.	6	1	99
5.	19	2	31	21.	17	2	78	37.	2	0	99
6.	23	2	33	22.	28	3	80	38.	2	0	99
7.	27	3	35	23.	16	1	82	39.	1	0	99
8.	23	2	38	24.	17	2	84	42.	2	0	99
9.	24	2	40	25.	25	2	86	45.	1	0	99
10.	43	4	44	26.	17	2	87	47.	1	0	99
11.	25	2	46	27.	24	2	90	50.	1	0	100
12.	38	4	50	28.	13	1	91	54.	1	0	100
13.	31	3	53	29.	16	1	92	69.	1	0	100
14.	51	5	57	30.	21	2	94	99.	3	0	100
15.	47	4	62	31.	18	2	96				

MEAN	13.013	MEDIAN	12.597
SKEWNESS	1.527	RANGE	98.999
VALID CASES	1070	MISSING CASES	153

TABLE 10

NUMBER OF STUDENTS IN LOWER 25% ACADEMICALLY

# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT
0.	373	35	35	11.	24	2	83	22.	10	1	97
1.	23	2	37	12.	27	3	85	23.	9	1	98
2.	44	4	41	13.	19	2	87	24.	6	1	98
3.	46	4	45	14.	12	1	88	25.	7	1	99
4.	69	6	52	15.	15	1	90	26.	1	0	99
5.	81	8	59	16.	14	1	91	27.	2	0	99
6.	64	6	65	17.	14	1	92	28.	4	0	100
7.	65	6	71	18.	17	2	94	29.	1	0	100
8.	32	3	74	19.	10	1	95	30.	1	0	100
9.	28	3	77	20.	4	0	95	31.	1	0	100
10.	36	3	80	21.	9	1	96	32.	2	0	100

MEAN	5.780	MEDIAN	4.210
SKEWNESS	1.348	RANGE	31.999
VALID CASES	1070	MISSING CASES	153

TABLE 11

NUMBER OF STUDENTS IN UPPER 25% OF SES

# OF STUDENTS	ABSOLUTE FREQUENCY	ABSOLUTE %	RELATIVE %
0.	776	63.5	73.2
1.	40	3.3	3.8
2.	54	4.4	5.1
3.	37	3.0	3.5
4.	35	2.9	3.3
5.	29	2.4	2.7
6.	16	1.3	1.5
7.	7	0.6	0.7
8.	8	0 7	0.8
9.	4	0.3	0.4
10.	15	1.2	1.4
11.	3	0.2	0.3
12.	5	0.4	0.5
13.	5	0.4	0.5
14.	5	0.4	0.5
15.	5	0.4	0.5
16.	5	0.4	0.5
18.	2	0.2	0.2
21.	1	0.1	0.1
23.	2	0.2	0.2
24.	1	0.1	0.1
25.	1	0.1	0.1
28.	1	0.1	0.1
30.	2	0.2	0.2
31.	1	0.1	0.1
	163	13.3	MISSING
	1223	100.0	100.0

MEAN	1.504	MEDIAN	0.184
SKEWNESS	3.917	RANGE	30.999
VALID CASES	1060	MISSING CASES	163

TABLE 12

NUMBER OF STUDENTS IN MID 50% OF SES

# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT
0.	90	8	8	17.	26	2	57	34.	6	1	97
1.	16	2	10	18.	31	3	60	35.	3	0	98
2.	39	4	14	19.	30	3	63	36.	6	1	98
3.	40	4	17	20.	44	4	67	37.	2	0	98
4.	35	3	21	21.	32	3	70	38.	2	0	99
5.	29	3	23	22.	39	4	74	39.	1	0	99
6.	32	3	26	23.	23	2	76	41.	1	0	99
7.	28	3	29	24.	25	2	78	42.	2	0	99
8.	26	2	32	25.	35	3	82	45.	1	0	99
9.	20	2	33	26.	24	2	84	47.	1	0	99
10.	39	4	37	27.	31	3	87	48.	1	0	99
11.	32	3	40	28.	21	2	89	50.	1	0	99
12.	30	3	43	29.	22	2	91	69.	1	0	100
13.	18	2	45	30.	31	3	94	91.	1	0	100
14.	37	3	48	31.	14	1	95	99.	4	0	100
15.	37	3	52	32.	10	1	96				
16.	32	3	55	33.	10	1	97				

MEAN	15.519	MEDIAN	15.027
SKEWNESS	1.761	RANGE	98.999
VALID CASES	1061	MISSING CASES	162

TABLE 13

NUMBER OF STUDENTS IN LOWER 25% OF SES

# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT
0.	375	35	35	11.	16	2	78	22.	11	1	98
1.	16	2	37	12.	38	4	82	23.	4	0	98
2.	42	4	41	13.	27	3	84	24.	6	1	99
3.	50	5	46	14.	20	2	86	25.	1	0	99
4.	30	3	48	15.	21	2	88	26.	1	0	99
5.	66	6	55	16.	21	2	90	27.	3	0	99
6.	47	4	59	17.	10	1	91	28.	4	0	100
7.	54	5	64	18.	12	1	92	29.	1	0	100
8.	41	4	68	19.	17	2	94	31.	1	0	100
9.	34	3	71	20.	20	2	96	35.	2	0	100
10.	56	5	77	21.	12	1	97	37.	1	0	100

MEAN	6.338	MEDIAN	4.758
SKEWNESS	1.109	RANGE	36.999
VALID CASES	1060	MISSING CASES	163

TABLE 14

NUMBER OF GIFTED STUDENTS

# OF GIFTED	ABSOLUTE EQUENCY	ABSOLUTE %	RELATIVE %
0.	877	71.7	78.1
1.	93	7.6	8.3
2.	55	4.5	4.9
3.	39	3.2	3.5
4.	9	0.7	0.8
5.	8	0.7	0.7
6.	9	0.7	0.8
7.	6	0.5	0.5
8.	6	0.5	0.5
11.	3	0.2	0.3
12.	1	0.1	0.1
14.	4	0.3	0.4
15.	4	0.3	0.4
16.	2	0.2	0.2
17.	1	0.1	0.1
18.	1	0.1	0.1
20.	2	0.2	0.2
23.	1	0.1	0.1
28.	1	0.1	0.1
31.	1	0.1	0.1
MISSING	100	8.2	MISSING
TOTAL	1223	100.0	100.0

MEAN	0.793	MEDIAN	0.141
SKEWNESS	5.988	RANGE	30.999
VALID CASES	1123	MISSING CASES	100

TABLE 15

NUMBER OF LEARNING DISABLED STUDENTS IN CLASSROOM

# OF LD STUDENTS	ABSOLUTE FREQUENCY	ABSOLUTE %	RELATIVE %
0.	612	50.0	54.4
1.	151	12.3	13.4
2.	106	8.7	9.4
3.	63	5.2	5.6
4.	49	4.0	4.4
5.	33	2.7	2.9
6.	27	2.2	2.4
7.	21	1.7	1.9
8.	11	0.9	1.0
9.	7	0.6	0.6
10.	8	0.7	0.7
11.	9	0.7	0.8
12.	5	0.4	0.4
13.	5	0.4	0.4
14.	1	0.1	0.1
15.	3	0.2	0.3
16.	3	0.2	0.3
18.	6	0.5	0.5
19.	2	0.2	0.2
22.	2	0.2	0.2
27.	1	0.1	0.1
MISSING	98	8.0	MISSING
TOTAL	1223	100.0	100.0

MEAN	1.779	MEDIAN	0.420
SKEWNESS	3.052	RANGE	26.999
VALID CASES	1125	MISSING CASES	98

TABLE 16

NUMBER OF ENGLISH SPEAKERS

# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT	# STUDT	FREQ	ADJ PCT	CUM PCT
0.	12	1	1	19.	36	3	30	38.	2	0	97
1.	1	0	1	20.	51	4	34	39.	3	0	98
2.	17	1	3	21.	48	4	39	40.	3	0	98
3.	5	0	3	22.	62	5	44	41.	2	0	98
4.	14	1	4	23.	53	5	49	42.	3	0	98
5.	18	2	6	24.	60	5	54	43.	1	0	99
6.	18	2	7	25.	65	6	59	45.	3	0	99
7.	18	2	9	26.	55	5	64	47.	1	0	99
8.	14	1	10	27.	72	6	70	48.	1	0	99
9.	14	1	11	28.	50	4	75	49.	1	0	99
10.	13	1	12	29.	54	5	79	50.	2	0	99
11.	18	2	14	30.	64	6	85	51.	2	0	99
12.	17	1	15	31.	36	3	88	53.	1	0	99
13.	12	1	16	32.	33	3	91	57.	1	0	100
14.	28	2	19	33.	29	2	93	60.	1	0	100
15.	26	2	21	34.	22	2	95	69.	1	0	100
16.	15	1	22	35.	7	1	96	99.	3	0	100
17.	25	2	25	36.	11	1	97				
18.	28	2	27	37.	8	1	97				

MEAN	22.691	MEDIAN	23.784
SKEWNESS	0.987	RANGE	98.999
VALID CASES	1160	MISSING CASES	63

TABLE 17

NUMBER OF NON-STANDARD ENGLISH SPEAKERS

# STUDENTS	ABSOLUTE FREQUENCY	ABSOLUTE %	RELATIVE %
0.	1019	83.3	88.1
1.	38	3.1	3.3
2.	27	2.2	2.3
3.	20	1.6	1.7
4.	13	1.1	1.1
5.	7	0.6	0.6
6.	6	0.5	0.5
7.	5	0.4	0.4
8.	2	0.2	0.2
9.	3	0.2	0.3
10.	5	0.4	0.4
11.	3	0.2	0.3
12.	3	0.2	0.3
13.	1	0.1	0.1
15.	1	0.1	0.1
17.	2	0.2	0.2
18.	1	0.1	0.1
MISSING	67	5.5	MISSING
TOTAL	1223	100.0	100.0

MEAN	0.478	MEDIAN	0.068
SKEWNESS	5.443	RANGE	17.999
VALID CASES	1156	MISSING CASES	67

TABLE 18

NUMBER OF SPEAKERS OF ENGLISH AS A SECOND LANGUAGE

# STUDENTS	ABSOLUTE FREQUENCY	ABSOLUTE %	RELATIVE %
0.	971	79.4	84.0
1.	88	7.2	7.6
2.	40	3.3	3.5
3.	21	1.7	1.8
4.	8	0.7	0.7
5.	10	0.8	0.9
6.	1	0.1	0.1
7.	2	0.2	0.2
8.	1	0.1	0.1
9.	1	0.1	0.1
10.	2	0.2	0.2
11.	2	0.2	0.2
12.	1	0.1	0.1
13.	1	0.1	0.1
14.	2	0.2	0.2
16.	1	0.1	0.1
17.	2	0.2	0.2
18.	2	0.2	0.2
MISSING	67	5.5	MISSING
TOTAL	1223	100.0	100.0

MEAN	0.460	MEDIAN	0.096
SKEWNESS	6.577	RANGE	17.999
VALID CASES	1156	MISSING CASES	67

TABLE 19

NUMBER OF STUDENTS WHO SPEAK ONLY A FOREIGN LANGUAGE

# STUDENTS	ABSOLUTE FREQUENCY	ABSOLUTE %	RELATIVE %
0	1132	92.6	97.9
1	20	1.6	1.7
2	3	0.2	0.3
3	1	0.1	0.1
MISSING	67	5.5	MISSING
TOTAL	1223	100.0	100.0

MEAN	0.026	MEDIAN	0.012
SKEWNESS	9.055	RANGE	2.999
VALID CASES	1156	MISSING CASES	67

TABLE 20

NUMBER OF YEARS TEACHING

# YRS	FREQ	ADJ PCT	CUM PCT	# YRS	FREQ	ADJ PCT	CUM PCT	# YRS	FREQ	ADJ PCT	CUM PCT
0.	30	3	3	13.	53	4	61	26.	14	1	95
1.	31	3	5	14.	74	6	67	27.	10	1	96
2.	52	4	10	15.	51	4	71	28.	6	1	97
3.	46	4	13	16.	42	4	75	29.	9	1	97
4.	53	4	18	17.	37	3	78	30.	11	1	98
5.	43	4	21	18.	33	3	81	31.	2	0	99
6.	50	4	26	19.	18	2	82	32.	6	1	99
7.	60	5	31	20.	31	3	85	33.	1	0	99
8.	55	5	35	21.	26	2	87	34.	3	0	99
9.	45	4	39	22.	25	2	89	35.	3	0	100
10.	77	6	46	23.	19	2	91	36.	1	0	100
11.	61	5	51	24.	16	1	92	37.	3	0	100
12.	69	6	57	25.	23	2	94				

MEAN	12.117	MEDIAN	11.361
SKEWNESS	0.618	RANGE	37.000
VALID CASES	1189	MISSING CASES	34

TABLE 21

INSTRUCTIONAL METHOD REVISED TO ONLY THREE CATEGORIES

REVISED INSTRUCTIONAL METHOD	ABSOLUTE FREQUENCY	ABSOLUTE %	RELATIVE %
LECTURE ONLY	69	5.7	6.3
INTERACTION OF ANY TYPE	908	75.3	82.8
INDEPENDENT WORK	119	9.9	10.9
MISSING	110	9.1	MISSING
	1206	100.0	100.0

VALID CASES	1096	MISSING CASES	110

Appendix C

Standardizing Observation Scores

A primary concern in attempting to set norms for an instrument developed for use in a variety of school contexts is the process of standardizing scores. Probably the best method for developing such scores is the one followed by the vast majority of achievement test developers. Stratified samples of possible test takers respond to the test, and their absolute raw scores are standardized to scale scores through the use of linear transformations. In this situation, an individual's raw score can be related to the scores of all students in the norm group to specify his/her relative ranking, which is generally represented by either a Z score, T score, stanine or percentile rank.

Techniques such as those described above do not take into account the effects of observer/rater bias so commonly found in observation instruments. Such biases can produce different scores for the same cluster of behaviors if observed by different observers. The use of two or more observers considerably reduces the chance of such bias (Micceri, 1983), however, another method for reducing bias is the use of scores standardized across observers. This method should assure that if ten teachers are observed by ten different teams of observers, they will be ranked in approximately the same order across all ten teams. The reliability study conducted on the summative instrument analyzed data standardized across teachers and teams of observers.

The general procedure for standardizing across observers requires the standardization of each item in the following fashion:

1. Each observer observes at least ten teachers in different subject areas and preferably different schools. This should assure a variety of teacher quality levels for the behaviors of interest.

2. Each item frequency is then given a percentile rank relative to the same item for the other nine teachers observed. For example, if item 5 is marked 10 times for teacher 1, 9 for 2, 8 for three, etc., then teacher 1 would receive a percentile rank of .90 and teacher 10 a rank of .00. These percentile ranks shown above are based on the percentage of cases

falling below a specific score. Other common methods of creating percentile ranks are: the percentage of cases falling at or below a score, and the percentage of cases falling below plus one-half of all cases falling at the score. Each creates a slightly different percentile rank for a specific score. The majority of achievement test developers (approximately 90%), use the first method mentioned, due to their large samples. In our study, we follow this method. Computer algorithms, in general use either the second or third method, or some similar alternative due to the small number of cases often used in percentile rank determination.

a) The process of transforming raw item scores into percentile scores reduces the detrimental effects of extremely long tailed distribtions on resulting analyses. Item distributions tend to closely resemble exponential distributions, having extremely long straggling tails.

3. The percentiles are then transformed into expected normal scores using the Blom procedure within SAS (1982). Each item score is transformed to the expected normal score given its relationship with each other item score for the observer (or team) in question.

a) Normal scores are approximations to the expected order statistics for the normal (function). They are computed from ranks, with the resulting variables appearing normally distributed. Thus, if there were 100 ranks, and an individual achieved a rank of 5, his expected normal score would by approximately 1.65. In this study, normal scores were computed using SAS version 1982; the inverse cumulative normal function, the individuals rank, and the number of non-missing observations for the ranking variable (Blom, 1958).

b) For this study the normal scores were standardized to a mean of zero and a standard deviation of 1. They were then transformed to a mean of 5 and a standard deviation of 1. These scores for each item are then summed and a total score for each teacher is obtained. Twenty positive items, each having a mean of 5, results in a total mean score of 100, or 20 X 5. An individual who scored one standard deviation below the mean for teachers observed by his/her team on each item would receive a score of 80 ((100 + (20 X -1)).

The following is an example of this standardizing process applied to a single item observed by two observers having widely differing time senses across ten teachers. One observer tends to mark numerous behaviors, while the other notes very few for the same teacher on the same item. If raw scores were the basis for comparison, this difference in observer preference would make a tremendous difference in each teacher's score. The use of items standardized across observers, however, results in similar rankings for each teacher by both observers. Two main assumptions underlying this standardizing procedure are:

1. A small group of teachers (20 or less) will adequately represent the entire population of teachers. Thus, if an observer observes at least 10 teachers, he/she will observe a wide enough range of quality to allow for discrimination of teacher effectiveness.

2. An observer will tend to pace his/her observations on a specific item the same across various teachers.

As the Table shows, if these teachers were to be compared on this item by raw scores, observer 2 would rate each teacher considerably higher than observer 1. The standardized score for each, however, retains almost the same score for each observer, leaving in a slight variety that represents individual observer bias toward specific manifestations of generalized behaviors. When standardized scores are summed over 20 items, these slight variations noted tend to even out.

Thus, each teacher is rated on a scale relative to the other teachers observed by each observer (or team or observers). In order to assure a range of scores for a team, it is necessary that each observer view a number of teachers in different subject areas.

This method of standardization was used in the original analysis of the normative study data (as well as in the original analysis of the reliability study data - 1982/83). Although useful for reducing observer bias, this technique was deemed inappropriate for use in an application such as merit pay for the following reasons:

1. The assumption that a small number of teachers represents the entire population of teachers is inappropriate for the development of absolute scales.

2. It is impossible to relate standardized scores such as these to a normative population.

TABLE 22

Two Observers Observing the Same Ten Teachers

Teacher name	Raw Score		Percentile Rank		Normal Score	
	observer #1	observer #2	obs #1	obs #2	obs #1	obs #2
1 - Smith	1	15	13	13	- 1.27	- 1.27
2 - Jones	2	19	40	31	- 0.26	- 0.52
3 - Hayes	3	28	68	72	0.52	0.66
4 - Regin	2	20	40	49	- 0.26	- 0.00
5 - Borg	2	20	40	49	- 0.26	- 0.00
6 - Jay	5	45	90	90	1.55	1.55
7 - Bird	3	27	68	63	0.52	0.38
8 - Haven	4	36	81	81	1.00	1.00
9 - Kraven	1	15	13	13	- 1.27	- 1.27
10- Raven	2	19	40	31	- 0.26	- 0.52

The Quartile based scaling technique described in Appendix D was used instead of this technique.

Appendix D

Quartile Based Scoring Method

The data base used in this study was divided into two randomly selected groups (n=670, and n=553). To develop a scaling technique for these data, it was necessary to consider the following factors:

1. Item distributions tend to be "J" shaped, almost approximating the exponential. This means they have extremely long, straggled tails. It is necessary to attempt to reduce the effects of these long tails on the resulting statistical analysis when considering scaling. For this reason, normal standardized scores were considered as a scaling technique. However, they were not appropriate for this application, for the reasons noted in Appendix C.

2. Although the underlying assumptions of this instrument dictate that more behaviors among more indicators result in better instruction, this does not necessarily mean that "more is better" for any specific behavior. In some cases, "more" may well be "worse". For this reason, it is necessary to limit the influence on a total score resulting from any specific item on the instrument. This avoids the problem of overweighting that results when more frequently occurring behaviors are weighted equally against less frequently occurring behaviors by frequency or percentage.

In view of the above considerations, it was decided to use the semi-interquartile range as a basis for scaling items. Thus, individuals who exhibit a frequency of behaviors equal to or above the top 25% as related to the normative data base receive a score of 3; those who fall between the 25th and 75th receive a score of 2; and those who fall into the bottom 25th percent receive a score of 1 point. This scaling by item may be related to the merit concept: those in the top 25% are considered "Better", those in the middle 50%, "Competent", and those in the bottom 25%, "Questionable".

It is not expected that each teacher score in the top or even the middle 25% on every item, as each lesson and topic may require a slightly different mix of behaviors for effec-

tiveness. However, the total distribution of item frequencies across all behaviors will indicate the activity of a teacher during a lesson. By summing scores, both positive (effective) and negative (ineffective), it is possible to relate a teacher's behaviors to the entire norming sample without any one item overly influencing a teacher's standing.

Techniques Applied in Developing Scale

The following consideration was of primary import during the development of a scale basis:

1. One certified trainer may pace him/herself at a totally different rate than another certified trainer. Thus it is possible for one observer to note substantially more behaviors than another. This suggests that **extreme observations** should be excluded from the norming basis for scaling.

To control for the possible effects of this factor, all observers whose average item frequencies for 50% of the items were more than twice the average for all observers were excluded from analysis in setting standards for scaling.

Scaling standards were based on fractions for each item. All levels of behavior were taken to the fraction represented by the specific quartile point (25%,75%) for subsample 1 (n=670). These results were then compared to the results from the second subsample, and found to correspond closely. In addition, the scores resulting from the first subsample were correlated with normal standardized scores on the second subsample and found to correlate r = .74. Thus, the scores resulting from this technique were in substantial agreement with those obtained using a standardization process across observers (described in Appendix C). The two subsamples were then combined to obtain final standards.

Notes

1. One item, "Begins Instruction Promptly" was treated as a categorical variable as opposed to a multi-point variable (3 points), as a teacher either begins promptly, or fails to begin promptly. This behavior usually occurs only once in a lesson.

2. Fractional scores are used as standards for the following reasons:

a) In the norming study each item score is a fraction as a result of controlling for the length of the observations.

b) In the field, as in the reliability study, teachers will be observed by at least three observers, and each item score will be the mean of three observations, which, in most cases will be a fraction. Thus, if two out of three observers record a teacher as "beginning instruction promptly", this teacher will score a point for this behavior. If, however, only a single observer records the teacher on this behavior, the scale assumes that it is not a consistent element in this teacher's lessons.

Appendix E

Total Scores Compared Across Frame Factors

Introduction

This appendix lists tables not found in the narrative. It includes tables showing positive (effective) scores adjusted for Instructional Method and unadjusted negative (ineffective) scores. All scores are compared across important Frame Factors such as general subject matter area for each of the two norm groups separately.

For the total groups the following positive means apply:

	Mean	Std Dev	N
Elementary	38.55	6.07	468
Post Elementary	35.50	6.70	721
totals	36.70	6.63	1189

and the following unadjusted negative mean total scores:

	Mean	Std Dev	N
Elementary	35.97	1.84	468
Post-Elementary	36.41	2.22	721
	36.23	2.09	1189

TABLE 23

Total Positive Scores Across Subjects by Norm Groups

Subject Area	Elementary		Post Elementary	
	Mean	n	Mean	n
Language Arts	38.52	240	35.83	170
Math	37.85	61	35.05	110
Science	36.53	19	35.36	100
Social Sciences	39.18	26	35.01	93
Practical Arts	37.91	50	35.23	181
Special Education	39.85	37	36.03	32

TABLE 24

Total Positive Scores Across Arts Subjects by Norm Groups

Subject Area	Elementary		Post Elementary	
	Mean	n	Mean	n
Art, Drama, Speech	38.24	7	33.13	11
Physical Education	36.48	16	33.03	27
Music	35.18	11	36.36	17
Agriculture	-----	--	37.10	7
Business/Marketing	40.50	2	36.34	19
Clerical/Sect/Data	37.51	6	34.98	26
Industrial/DCT/Shop	-----	--	35.78	18
Home Ec./Crafts	-----	--	32.93	29
Other	44.00	3	37.50	3

TABLE 25

Total Positive Scores Across Classroom Types by Norm Groups

Classroom Type	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
Self Contained	39.05	339	35.66	625
Pod or Open Class	36.68	70	35.41	14
PE Court or Field	35.12	14	35.14	20
Lab	37.50	2	36.31	16
Resource Room	40.34	15	35.20	7
Departmental	40.97	9	32.59	7
Media/Clinic	38.40	5	33.50	2
Other	35.33	3	30.57	3

TABLE 26

Total Positive Scores Across Ethnic Groups by Norm Groups

Ethnic Group	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
White-Non Hispanic	38.68	376	35.72	635
Black-Non Hispanic	38.28	65	33.04	58
Hispanic	38.00	14	37.36	11
Other	35.00	3	35.00	3

TABLE 27

Total Positive Scores Across Teacher's Degree by Norm Groups

Highest Degree Attained	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
Non-Degree	-----	--	35.43	8
BA.,BS.,BFA.	38.49	290	35.63	368
MA.,MS.,MED.	38.73	165	35.39	289
Ed.S.,Ed.D.,Ph.D.	37.63	8	34.85	30

TABLE 28

Total Positive Scores Across Subjects by Norm Groups

Years Experience	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
One Year or Less	40.02	59	36.36	33
2 to 4 years	39.21	66	36.41	83
5 to 9 years	38.80	104	35.59	141
10 to 14 years	38.67	125	35.68	199
15 or more years	37.83	141	34.98	241

TABLE 29

Total Positive Scores by SES Characteristics by Norm Groups

Percentage of Low SES Students	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
20% or less low SES	38.11	180	35.42	368
20% to 39% low SES	38.77	128	35.82	87
40% to 59% low SES	40.23	27	33.82	37
60% to 79% low SES	41.70	11	37.97	18
80% or more low SES	38.97	96	34.37	105

TABLE 30

Total Positive Scores Across Sex by Norm Groups

Sex	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
Male	37.76	53	35.27	285
Female	38.67	410	35.65	429

TABLE 31

Total Negative Scores Across Subjects by Norm Groups

Subject Area	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
Language Arts	35.92	240	36.60	170
Math	35.98	61	36.16	110
Science	35.89	19	36.43	100
Social Sciences	36.35	26	36.02	93
Practical Arts	36.14	50	36.73	181
Special Education	35.84	37	35.63	32

TABLE 32

Total Negative Scores Across Arts Subjects by Norm Groups

Subject Area	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
Art, Drama, Speech	36.28	7	37.45	11
Physical Education	36.44	16	36.85	27
Music	36.00	11	36.18	17
Agriculture	-----	--	35.43	7
Business/Marketing	36.50	2	36.79	19
Clerical/Sect/Data	35.83	6	37.08	26
Industrial/DCT/Shop	-----	--	37.16	18
Home Ec./Crafts	-----	--	36.48	29
Other	37.00	3	36.50	3

TABLE 33

Total Negative Scores Across Classroom Types by Norm Groups

Classroom Type	Elementary		Post Elementary	
	Mean	n	Mean	n
Self Contained	35.92	339	36.37	625
Pod or Open Class	36.29	70	37.64	14
PE Court or Field	36.57	14	36.60	20
Lab	35.00	2	36.63	16
Resource Room	35.93	15	37.14	7
Departmental	36.22	9	36.00	7
Media/Clinic	36.00	5	36.00	2
Other	36.00	3	37.00	3

TABLE 34

Total Negative Scores Across Ethnic Groups by Norm Groups

Ethnic Group	Elementary		Post Elementary	
	Mean	n	Mean	n
White-Non Hispanic	36.05	376	36.51	635
Black-Non Hispanic	35.75	65	35.62	58
Hispanic	36.07	14	36.00	11
Other	33.00	3	36.67	3

TABLE 35

Total Negative Scores Across Highest Degree by Norm Groups

Highest Degree Attained	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
Non-Degree	-----	--	36.25	8
BA.,BS.,BFA.	35.90	290	36.34	368
MA.,MS. MED.	36.07	165	36.42	289
Ed.S.,Ed.D.,Ph.D.	36.63	8	36.63	30

TABLE 36

Total Negative Scores Across Experience by Norm Groups

Years Experience	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
One Year or Less	35.62	59	36.24	33
2 to 4 years	36.00	66	36.39	83
5 to 9 years	36.09	104	36.45	141
10 to 14 years	35.99	125	36.44	199
15 or more years	35.92	141	36.29	241

TABLE 37

Total Negative Scores by SES Characteristics by Norm Groups

Percentage of Low SES Students	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
20% or less low SES	36.18	180	36.42	368
20% to 39% low SES	35.92	128	35.84	87
40% to 59% low SES	35.37	27	36.16	37
60% to 79% low SES	35.09	11	35.94	18
80% or more low SES	35.92	96	35.71	105

TABLE 38

Total Negative Scores Across Sex by Norm Groups

Sex	Elementary Mean	Elementary n	Post Elementary Mean	Post Elementary n
Male	35.96	53	36.11	285
Female	35.97	410	36.63	429

Appendix F

Procedure Used in Developing Adjustments

Adjustments for a three category instructional method variable including: lecture only; any type of interaction, disussion, recitation, and independent work, were developed by subtracting the scores obtained by teachers using the two lower scoring instructional methods (lecture and independent) from the mean for random subsample 1, and then testing these scores on the second subsample by adding the difference to each of the lower scoring groups in the second subsample. Due to the small number of cases for each group (#1 n=69, #3 n=119), the stability of these tests is questionable, however, the results were reasonably close (within the 2.15 criterion) for both groups. Since 85% of lecture only and independent seatwork occurred in post-elementary classes, only norm group 2 (post-elementary) was used for this purpose.

Following this, the two samples were combined and differences subtracted to obtain adjustment factors for both of these instructional methods from the total data base.

As this instrument is essentially an interaction based measure of teaching behavior, these adjustments fulfill three criteria:

1. significance,
2. meaningfulness, and
3. logic.

When only lecture is used, or independent seatwork, there is far less opportunity for teacher interaction than with the interaction technique. It is only to be expected that those groups would score lower.

Basis and Scores Used in Adjustment

The following scores were obtained for the entire population of post elementary classes:

TABLE 39

UNADJUSTED TOTAL SCORES ACROSS INSTRUCTIONAL METHOD

FOR POST-ELEMENTARY SCHOOL TEACHERS

INSTRUCTIONAL METHOD	MEAN	STD DEV	N
LECTURE ONLY	32.92	5.89	64
INTERACTION/DISCUSSION	36.13	6.70	487
INDEPENDENT SEATWORK	30.25	6.46	95
Total Group	34.96	6.86	646

Based upon these scores, adjustment upward for Lecture Only resulted in a score addition of 2.03 points, and for Independent seatwork one of 4.70.

Appendix G

Results of Positive Scores

Results of Regression Analysis

The multiple r's obtained through regression anlaysis across 46 different dependent variables using 18 independent variables were uniformly insignificant. The highest multiple r obtained from this plethora of analyses was .36 (positive item #8 - gives specific academic praise), and the largest contributor (simple r = .25) was the grade level of the classroom (treated as a continuous variable). Note that this factor - Grade Level - is used as the basis for forming two separate norm groups, which adequately controls for its effects on teacher scores.

General Results of Anovas

Since regression indicated that individual item scores were not adversely influenced by Frame Factors, the major portion of Anova analysis was conducted on total scores (negative and positive), and domain scores. Each item was included in an analysis with each independent categorical variable. Again, the meaningfulness criterion (2.15) resulted in almost no important relationships.

Oneways.

Using oneway ANOVAs on subgroup 2, (both parametric and non-parametric), the following variables proved highly significant across levels in relation to total positive scores (listed in order of significance): Instructional Method (IM), Grade Level (Elementary, middle or secondary), Class Grouping, Subject Matter area, and Classroom Type if not self-contained. When cross-validated upon Group #1 (n=670) using a factorial unweighted means analysis (Blair and Higgens, 1978; Kirk, 1968), only Grade Level retained its significant and seemingly meaningful main effect when crossed with instructional method. The data contained in this study violate the major assumption underlying most unbalanced factorial ANOVAs, random assignment to groups. Lacking this criteria, it is best to use the conservative random effects unweighted means model.

Results following Adjustments for Instructional Method

Oneway Analyses of Variance were run across all important categorical Frame Factors across the total data base. As a result of these analyses, the following variables were found significant at the (.01) level:

1. Positive Scores:

   a) Grade Level, and

   b) Subject Matter Area.

2. Negative Scores:

   a) Grade Level,

   b) Sex, and

   c) Ethnicity.

3. All differences disappeared when analyzed in separate norm groups (elementary and post-elementary), as much of the difference between groups in all the above relates directly to the relative number of teachers in a group that will be found in elementary classes (for example, note that only 16% of male teachers were found in elementary classes, while 39% of all teachers in our sample taught in elementary schools).

Section II, Appendices A-B

DOMAIN--Formative Instrument

6.0 Testing: Student Preparation, Administration, Feedback

6.0 TESTING: STUDENT PREPARATION, ADMINISTRATION, FEEDBACK

Definition: This domain = development and maintenance of an environment in which students can validly demonstrate their knowledge, skills, etc., and receive adequate information about the quality of their test performance.

This domain includes the following named concepts:

6.1 Preparation for Testing

6.2 Test Administration

6.3 Formative Feedback

TESTING

Overview

Testing is an integral part of instruction. Test data are essential for the teacher in making day to day decisions and for the learner in understanding his capacities, achievement, and learning difficulties.

Because tests can have a decided influence upon students, they should not be taken lightly or given haphazardly. To assess student progress, the teacher must have knowledge of how to plan and construction tests, treatment of which is found in the Handbook on Measurement and Evaluation. However, this is not sufficient. The teacher must also know how to administer tests, how to prepare students in the knowledge and skills of test-taking, and how to provide feedback on test performance. These conditions of successful testing are treated in this Domain.

Research studies indicate that students do their best when they are motivated and when they are aware of, and apply, test-taking skills. It is therefore important that teachers motivate students to put forth their best efforts and that they prepare them in the skills of test-taking. Furthermore, cheating has become a major problem in today's schools. While not all cheating can be attributed to inadequate physical setting and poor classroom supervision, these factors can be important in the reduction of cheating.

The use that teachers make of test data is a significant aspect of instruction. Teachers who keep the results of test to themselves reduce the motivation of their students and deny to them information about their own successes and failures and what they need to do to improve their achievement. Frequent testing and feedback to the learner can be, and typically is, an effective way to increase achievement. Frequent feedback is particularly beneficial to learners from poor educational backgrounds and to students with little motivation.

This Domain deals with the foregoing components of testing. It sets forth research knowledge, the application of which can result in reduced student fear of examinations, improved test motivation and test-taking skills, reduced cheating, and a feeling by students that they have been fairly judged.

6.1 Concept: Preparation for Testing

Definition: Preparation for testing = teacher verbal behavior and actions prior to testing that ready the students for tests.

Indicators

6.1.1 Orients Students to Test

Definition: Orients students to test = teacher's verbal behavior that informs students of the purpose of the test, how the test results will be used, and how the results are relevant to them personally.

Example (Pos): Mr. Chen explains to his typing class that every day they will have timed tests. The test scores are to be recorded on individual charts to help them improve their typing rate. The daily scores will not be part of their class grade, but he will look at the scores to guide their improvement. He assures them they can all increase their typing rate with practice.

Example (Pos): Mr. Benson is administering the Stanford Diagnostic Mathematics Test to his six grade class. Mr. Benson states: This test is designed to help me tell which math concepts we have been studying that you do not understand. I will use the results to place you in groups so that I can better help you to master those concepts which are not yet clear to you. The results will not be used in determining your grade, but rather to help us be sure that when you go on to 7th grade math you do know 6th grade math!

Example (Neg): Ms. Jensen tells her class she is going to give them an achievement test tomorrow to find out who should be making the "A's" and "B's" and who should get "F's." Furthermore, if anyone doesn't "measure up" on the test she might confer with that student's parents.

6.1.2 Specifies Test Content

Definition: Specifies test content = teacher tells the class what knowledge is important for them to study and suggests references or guides.

Example (Pos): A few days before the home economics test, Ms. Hernandez reviews the food groups covered in class. She gives the students a study guide containing study questions to take home. She specifies what chapters and topics will be covered on the examination, and tells the class that the problems on the test will be similar to those they have been working on.

Example (Neg): Ms. Wilson tells the students the test they are to take tomorrow includes all kinds of questions about everything they should have learned, so they had better take their books home tonight and study hard.

6.1.3 Instructs Students in Test-Taking

Definition: Instructs students in test-taking = teacher behavior which increases the students' abilities to utilize the characteristics and format of a test and/or test-taking situation to improve their chances of making a higher score.

Example (Pos): In an English class, Mr. Graff instructs his students in how to take a test. He tells them to read the directions and begin work immediately; to mark the items they know, omitting the hard items; to then go back to the hard ones and by progressive elimination, choose the best answer; to check all answers when test is finished. He explains how to use deductive reasoning to choose between options and provides practice in this process.

Example (Neg): Mr. Moto tells his class that there is no way to answer the test questions without mastery of the subject. Study hard. If you don't know the answer, then do the best you can.

Principles

If the teacher informs students of the purpose of the test, how the results will be used and why the results are relevant to them personally, then their test performance will likely improve.

If the teacher instructs students in utilizing the characteristics and formats of tests and/or test-taking situations to receive higher scores, then the tests will be more valid measures of what they are intended to measure.

Supportive Evidence

The perceived importance of a test appears to affect student performance on the test. Anxiety can be a related factor. See Supportive Evidence under 6.2, Test Administration.

Yamamoto and Dizney (1965) found that elementary and secondary students, particularly girls, did better on the average when a test was called an "intelligence" test than they did when the test was called an "achievement" or "routine" test. Naughton (1968) informed a third of his 10th grade class as to how their scores on the Differential Aptitudes Test were to be used and stressed the importance of maximum effort on the test.

The fact that teacher behavior before a test can influence student performance was demonstrated in a study by Gorden and Durea (1948). Scores for a group of students who were discouraged by the examiner were significantly depressed, one-third of a standard deviation on the average, compared to scores of a group that was not discouraged.

The effect of expectancy is supported in a study by Osler (1954). She divided 15 high school classrooms into three groups. One group was informed that they were among the lowest 10% on a pretest in arithmetic, another group that they were the most successful, and a third group given no information. When a test was administered to all groups, she found that students in the success and control groups did significantly better than students in the failure group. Apparently when students are led to perceive themselves as failures, they function that way.

Teachers can help students get ready for a test by giving them specific information about the content to be covered and the type of test to be given. High school students given training in notetaking which emphasized distinguishing between subordinate and superordinate information, abbreviating words, paraphrasing and using an outline format, and who expected a multiple choice test, did significantly better than those who did not receive training; furthermore, they took more efficient notes. In the group who expected an essay test, subjects with no training scored higher than trained students, but among those trained, the students who expected a multiple choice test were better prepared and scored higher on the essay test than those who expected an essay test (Carrier and Titus, 1981).

Test anxiety was reduced and achievement was higher for a group of high anxiety college students who were given study guides for exam preparation. This was true for high, medium and low ability levels according to Szafran (1981).

Another way teachers can help students prepare for tests is to teach them how to take a test. Instruction might include format, content clues, use of test time strategy, etc., and could include practice particularly if separate answer sheets are to be used. Test-taking skills are widely recognized as a source of variance in test scores and as a possible depressor of test validity. Testwiseness permits some students to score higher on classroom and standardized tests than other students who lack test-taking skills, although the students may be equally knowledgable of the subject matter.

Oakland (1972) identified prerequisite abilities needed for taking the Metropolitan Readiness Test. Children from Head Start classes were randomly divided into experimental and control groups. The experimental group was given instruction on these skills (not on the content of the test) during 30-minute periods twice a week for six weeks. The experimental group scored significantly higher on the MPT immediately following the instruction but lost the advantage over the summer months for a second post-test given in September.

Kalechstein, Kalechstein and Docter (1981) report the effects of instruction in test-taking techniques including how to follow oral directions, how to mark responses, strategies for using test time, and guessing strategies, upon standardized reading test scores of second grade Black students. The experimental group which received the instruction performed significantly higher ($p < .01$) than the control group when administered the Stanford Achievement Test.

To determine the effect of test-taking skill training on performance on a school readiness test, an experimental group of Mexican-American kindergarten students received two twenty minute sessions of training. Dreisbach and Keogh (1982) found significantly higher scores for those students who received training, especially boys, regardless of whether the test was administered in Spanish or in English.

Slakter, Koehler, and Hampton (1970) showed that test-taking skills increased as children moved through the grades.

Extensions/Exceptions

One study indicates that emotional factors may play a part in test scores. Certainly the death of a loved one, the anticipation of a birthday celebration, or school experiences may be a positive or negative influence on test performance. Examinees' experiences immediately preceding a test affected their scores in studies by McCarthy (1944) and Reichenberg-Hackett (1953). In the McCarthy study, third and fourth grade children who wrote a composition on "The Best Thing That Ever Happened to Me" prior to a test averaged four to five points higher than their scores on the same test after writing on "The Worst Thing That Ever Happened to Me." Children in the Reichenberg-Hackett study showed significant improvement in their scores after a gratifying experience involving the successful solution of an interesting puzzle, followed by a reward of toys and candy. Improvement did not follow neutral or less gratifying experiences. Other benefits from the positive teacher behavior cited here include better attitude toward schoolwork, not just testing. Shannon (1980) found that student attitudes toward reading were more positive after taking a standardized reading achievement test when the students had been counseled on the purpose and nature of the testing before the test was administered.

It appears that teachers should be positive and approving in a test situation and convey the idea that there is an important task to be accomplished. Pupils should not be threatened with a test or warned of dire consequences if they fail.

Kirkland (1971) reviewed seven studies on the relation of expectation of success or failure to successful performance on tests. Expectation of success or failure on tests determined the amount of time students persisted in attempting to solve problems. Those who had high expectancy of success performed better than those who had a low expectancy.

Typically, teachers encourage students to skip over difficult items and return to them if time permits. Rindler (1980) reported on the correlation of skipping difficult items on a verbal aptitude test for low, middle and high ability students under varying time limits. For the high ability students, subjects who skipped items scored higher than those who did not. However, in the middle and low ability groups, students who skipped items performed more poorly than those who did not.

6.2 Concept: Test Administration

Definition: Test administation = teacher performance that provides favorable environmental conditions for test taking.

Indicators

6.2.1 Arranges Physical Setting

Definition: Arranges physical setting = teacher performance that sets up the physical arrangements (seating, and control of distractions and conditions of temperature, ventilation and lighting) favorable to student performance on examinations.

Example (Pos): The day of the announced examination for his class, Mr. Bloom entered his classroom early. Before the students entered he arranged the chairs to provide as much space as possible between students. He checked the temperature and ventilation and turned on all of the lights in the room. He placed a sign on the door which read "Testing in Progress."

Example (Neg): Ms. Vaughn enters the classroom and says: Okay, class. As you know you're going to have a test today on the last three chapters. We're going to have to move the desks. Please move your desk out of the circular formation and set them up in rows again. John, run down to the custodian's office and ask him if he can come up and do something with that flickering light this morning.

6.2.2 Make Positive Statements

Definition: Makes positive statements = teacher verbal expressions that set a warm and encouraging classroom atmosphere for the examination.

Example (Neg): Ms. Jacobs is faced with the task of administering the State Assessment tests to her 5th grade class. She stands before her class and states: Now you are going to be sorry you haven't been listening to me all year! If you don't pass this test, you won't go on to 6th grade next year. I bet I'll have half of you in my class again. Heaven forbid! Well, do the best you can. You'll have to work fast or you won't get through.

Example (Pos): Ms. Hudson's fifth grade class is scheduled for the State Assessment tests. Ms. Hudson greets her students cheerfully and states: You know that we have been working hard on what is in this test. I am sure you are well prepared and will do well on the test. Do your best and we will all be proud of the results!

6.2.3 Monitors Examination

Definition: Monitors examination = teacher behavior that tells students that he/she knows what they are doing.

Example (Neg): Mr. Woods is proud of his militarily run classroom, his knowledge of the subject matter, his difficult assignments, and his letting students know "their place." When the test booklets have been distributed, Julie raises her hand. MR. WOOD's responds: Julie, it is too late for questions now. After testing has begun, Ms. Allison from next door calls Mr. Wood from the room. They converse in the hall while their classes work on their tests.

Example (Pos): Mr. Page has reviewed the material for his first test and has answered questions to clarify students' understanding. He has invited students who do not feel ready for the test to see him after school for more help. When the students come into the classroom they find the chairs spaced as far apart as possible. Different forms of the test (on two colors of paper) are distributed. Mr. Page places a "Testing in Progress" sign on the door. He projects a friendly, pleasant manner, but at the same time communicates to students that he is alert to what they are doing.

Principles

If tests are administered in a physical setting which is comfortable and free of distractions and opportunities to cheat, by an examiner who is positive and encouraging, then students will have a fair chance to demonstrate what they know and the teacher will more likely obtain a valid measure of their achievement.

If the teacher is aware of the anxiety caused by tests in some students and refrains from using tests as a threat or from emphasizing the negative consequences of poor performance, then high anxious students will suffer less anxiety and perform better on tests.

Supportive Evidence

Walker, Sannito and Firetto (1970) reported that tension oriented instructions, such as telling examinees that everyone is expected to finish in the time allotted, had a beneficial effect on the performance of low anxious persons, but a deleterious effect on that of high anxious persons.

That debilitating anxiety can be induced by the examiner has been demonstrated in several studies. Wrightsman (1962) administered a timed measure of intelligence to two groups of college freshman under two conditions: 122 subjects were led to believe that performance on the test was very important and that the results would follow them throughout their college careers; another 112 subjects were led to believe the test was given only to collect norms. When the test was seen as important, anxious subjects scored significantly lower than non-anxious subjects. In the other condition, anxiety was unrelated to test performance.

For a test situation it is expected that students have comfortable seats of an appropriate size and adequate working space in a well-lighted, well-ventilated room with a minimum of outside noise. However, research on physical setting indicates that some latitude in these conditions can be tolerated without grave consequences. Traxler and Hilkert (1942) reported that high school students who used desks tended to obtain higher test scores than students using chairs with desk arms. However, Ingle and De Amico (1969) found that high school students who took tests with lap boards as a writing surface did no worse on a standarized achievement test than did students who had adequate writing surfaces.

Super, Braasch and Shay (1947) explored the effects of a variety of distractions on test performance of college students. The Otis Test of Mental Ability and two other tests were administered to thirty students in the distracted group and to twenty-six in the control group. The distracted group was subjected to a trumpeter playing in the next room, the examiner's friend bursting into the room and whispering loudly, "How long are you going to be here?", a loud argument outside the classroom door, and mistiming of the test with instructions for the test takers to proceed after the time clock bell rang. Comparison of the scores of the two groups of students found no statistically significant differences. The conclusion was drawn that commonly occurring distractions do not affect test results.

However, distractions may have an impact on scores of younger students. Trentham (1975) studied 72 sixth grade students randomly selected from three elementary schools. The students were randomly divided into experimental and control groups. The distractions used with the experimental group were an alarm bell, two kittens, blinking of the lights, a person entering the room to deliver a note, two talking teachers entering the room, and a radio playing the World Series for 30 seconds. The control group took the test undisturbed. The disturbed group scored significantly lower than the undisturbed control group. This finding supports the desirability of non-distracting testing conditions, at least for younger students.

Bushway and Nash (1977) reviewed the research on school cheating. They cite the studies of McQueen (1957) and Rogosin (1951) which conclude that situational factors in school are more important than behavioral characteristics of students as determinants of whether or not cheating occurs. These situational factors include the teacher's leaving the room during testing, the emotional tone of the classroom, and the moral climate of the school.

Fischer (1970) reported that 65% of the 135 boys and girls in a study in fourth, fifth, and sixth grades cheated under control conditions where students were given only usual test-taking instructions. In the experimental group 76% cheated under the conditions of appeal to honesty where the teacher explained the need for true measures of student's knowledge and effectiveness of her teaching techniques. Thirty-five percent cheated following a discussion of what students think about cheating and signing their names and writing reasons why they would not cheat on a test.

Vitro and Schoer (1972) found that the highest incidence of cheating occurred among students who were unlikely to do well on the test, who were unlikely to get caught, and to whom the particular test was very important.

Uhlig and Howes (1967) found that a large percentage of college students would cheat even in a non-pressure situation were the opportunity present.

Personality and teaching style of the teacher were found to be other situational determiners in cheating. Shirk and Hoffman (1961) found that teachers who are very authoritarian, know all the answers and infer that their students are inferior beings lead student to cheat more. Weldon (1966) found a positive relationship between amount of cheating and totalitarian procedures of teachers. Johnson and Klores (1968) and Steininger, Johnson and Kirts (1964) found that a dissatisfying classroom situation was judged by students as producing a greater amount of cheating.

Extensions/Exceptions

Several studies of test anxiety report 25-30% of the children studied as highly test anxious (Cotler and Palmer, 1970; Nottelman and Hill, 1977). Wine (1971) and Nottelman and Hill have suggested that the debilitating effects of test anxiety may have an attentional explanation. They report that high anxious students seem less able to attend to task-relevant cues, frequently look away from the task and perform more poorly. Wine suggested that high anxious students become preoccupied with (task-irrelevant) worry responses and do not devote enough attention to the test-taking task.

Culler and Holahan (1980) investigated the relation of test anxiety to academic performance in college freshman and the differences in study-related behaviors between high and low anxious individuals. They reported a significant decrement in grade point average associated with test anxiety (mean GPA for low-test anxious group was 2.86; for the high test anxious group, 2.51). High test anxious students were found to have poorer study skills.

Trentham (1979) did not find anxiety-producing instructions to sixth graders to be significant, but the high anxiety students did score significantly lower than medium and low anxiety students.

A comparison of SAT test scores of high school seniors by French (1962) did not indicate a significant effect of anxiety on validity. The only significant effect was high anxiety girls tended to do better on the mathematics test.

Green (1981) reported higher anxious students in statistics made more response changes. Correlations among anxiety, study skills and achievement test scores of college students reported by Kirkland and Hollandsworth (1979) show anxiety is negatively related to grade point average ($r = -.37$) and ACT scores ($r = -.31$). Effective Study Test scores and grade point average positively correlate (.37) ACT scores correlate .40 with Effective Study Test scores.

Smith (1969) reports that white second grade girls have significantly higher test anxiety than white second grade boys, but there was no difference in the test anxiety scores of black boys and girls. The differences in test anxiety scores between boys and girls increases through elementary school.

Bauer (1975) found that boys were more inhibited by anxiety producing instructions than girls. Clawson, Firment and Trower (1981) studied the effects of test anxiety on 150 ninth graders and 112 seventh graders in Northeast Florida. The students were administered the Test Attitude Inventory and the Comprehensive Test of Basic Skills. Mean scores for seventh and ninth grade blacks, in both language and math, centered on the 50th percentile. Whites scored at the 65th percentile and above on all measures. Blacks had significantly higher scores than whites in state, trait, and total test anxiety, by separate seventh and ninth grade groups, as well as in the combined group.

High state (transitory, situational) anxiety was found in 32% of the blacks and only 9.8% of the whites. On the other hand 19% of the whites had low trait (relatively stable characteristic tendency toward anxiety) test anxiety, compared to only 5.3% of blacks. Seven percent of all blacks and 14% of the whites were low in state test anxiety. Clawson et al contend that these findings must be considered in standardized testing, especially where results are used as a basis for graduation, promotion, and remediation.

The teacher may choose from alternative methods of presenting test items without significantly affecting test results. This can make it possible to provide test variety and meet the needs of exceptional children too. Bligh (1965) reviewed several studies which have shown that tests administered via television (Burr, 1963; Curtis and Kropp, 1961) or overhead projector (Schwarz, 1967) yield results comparable to those obtainable on conventionally administered tests, provided that speed is not a factor, and where other conditions are held constant. Burr (1963) studied scores sixth

graders earned on the social studies test of the Metropolitan Achievement Tests. He used three modes of item presentation: projected silence, audio-reinforced, and conventional administration. The experimental techniques used less time than the conventional approach, however, test scores were not significantly different for any test format. The ninth graders tested by Curtis and Kropp (1961) scored significantly higher on items involving word problems that were televised, but on quantitative items which gave computational examples, the normal administration produced significantly higher scores. Ninth graders had no difficulty with verbal abilities items that were televised. In fact, a significantly greater number of total items were responded to on the experimental administration.

In a study of in-class tests versus take-home exams, Marsh (1980) found significantly better results in all categories of the in-class tests. These students also indicated that they had studied significantly more. Kalish (1958) reported no significant difference in total errors between open-book and closed-book exams.

The effect of examiner's race in group tests has been studied by Jensen (1974). An elementary school with 60 percent white and 40 percent black students were given five tests. Twelve white and eight black examiners, administered the tests. The results of the five tests used showed the race of the examiner did not produce a significant impact on the scores of either black or white students.

Mishra (1980) found no significant effect of examiners' ethnicity on Anglo third grade students. Mexican-American students, however, scored significantly higher on vocabulary tests when they were administered by a Mexican-American examiner, but there were no significant differences on other tests.

Intelligence tests administered in English versus in Spanish or in English with an interpreter were composed by Swanson and Deblassie (1979). Spanish-speaking Mexican-American first and second grade children scored significantly higher on the Verbal tests when administered in English; however, in general, their scores were higher on the Performance tests with the Spanish administration. The authors discuss various influences which may have contributed to these findings such as the unfamiliarity with the use of Spanish in the scholastic setting leading to reticence on the part of the children. Other possible factors include the lack of direct Spanish counterparts to many English concepts and phrases, and the fact that facility in Spanish was an uncontrolled variable. The superiority of the Spanish administration of the Performance tests was hypothesized as attributable to decreased anxiety in the children.

Results of a non-verbal (perceptual/spatial) intelligence test with instructions given in English versus in Spanish were compared for bilingual Puerto-Rican sixth, seventh and eighth grade students (Anastasi and Cordova, 1953). No significant effects were found for language used in the instructions, order of the language used in the presentation or sex.

6.3 Concept: Formative Feedback

Definition: Formative feedback = teacher behavior that provides information to students about their test performance.

6.3.1. Reviews Tests

Definition: Reviews test = teacher uses test data to identify and correct errors in understanding.

Example (Pos): Ms. Thomas is aware of need for specific feedback concerning errors. The day after the test she scheduled a discussion session to go over the items on the test. She listened carefully to reasons students gave for alternative responses to questions. She identified areas of misunderstanding and clarified misconceptions.

Example (Neg): Mr. Zayer gives a test to his history class, saying he will return the tests the next day. The tests were actually returned a week later with grades and no comments. Students ask about the answers to certain questions. He tells them to refer to their textbook.

6.3.2 Comments Positively on Test Results

Definition: Comments positively on test results: teacher's verbal behavior that expresses approval of student response to test item(s) or to discussion of responses.

Example (Pos): Mr. Crawford returns test papers and calls attention to items most frequently missed. He asks some students to explain their answers and during the class discussion Mr. Crawford recognized different points of view and commented on the students' good thinking.

Example (Neg): Ms. Jenkins, on returning test papers, answers student explanation of their answers with such remarks as "How could you think that," "You must not have studied," "We discussed that thoroughly in class; you must not have listened."

Principle

If the teacher provides feedback to students on their tests, then motivation, learning, and retention will increase, and the teacher will be more able to adjust instruction to the needs of the classroom group.

Supportive Evidence

Students receive constant feedback from teachers, classmates and others in the school. The concern in this review is only with the effects of feedback from formative testing on later school achievement, retention and attitudes toward the subject matter.

Barringer and Gholson (1979) reviewed more than seventy studies on the effects of feedback on concept formation and concept identification in children. Three combinations of feedback were investigated: (a) feedback which follows both correct and incorrect responses (right-wrong); (b) feedback which follows correct responses, but does not follow incorrect responses (right-blank); and (c) feedback which follows incorrect responses, but does not follow correct responses (wrong-blank). They reported: that "In general, the research has been fairly consistent in two respects: (a) verbal and symbolic feedback produce more rapid acquisition than does tangible feedback, and (b) the right-blank combination is least efficient as a method for teaching children conceptual material. Wrong-blank has usually resulted in the most efficient performance during acquisition and the greatest resistance to extinction when all feedback was discontinued. The relative efficacy of right-wrong has been less consistent. In some cases, this combination has been similar, in effect, to wrong-blank; in others, it has been intermediate or similar to right-blank."

Reinforcing only correct responses is ineffective because children interpret the nonoccurrence of feedback as positive feedback. When no feedback follows incorrect responses they repeat their mistakes (Mims and Gholson, 1977).

Token economy has been effectively used to modify a variety of classroom social behaviors (aggression, social withdrawal, and inappropriate talking). Tangible feedback is less facilitative of conceptual learning, however, than is verbal or symbolic feedback. Some investigators (Mims and Gholson, 1977; Ratliff, 1972; Spence, 1970) have suggested that tangible feedback distracts the child's attention from relevant stimulus cues.

Hughes (1973) studied the effects of the quality of feedback on seventh grade students in 13 classrooms. The teacher planned the lesson in detail and memorized the presentation. Lesson content was held constant while teacher reactions to student responses were manipulated. The teacher either praised or reacted neutrally. Subsequent testing showed that positive teacher reactions facilitated student achievement significantly more than minimal (neutral) reactions.

Page (1958) demonstrated the importance of specificity of feedback. In 74 classes teachers administered an objective test to their students. A score and a grade was assigned to each paper and each paper was randomly assigned to one of three groups. Group I papers received no comment other than a score and a grade; Group II papers received general and encouraging comments, like "Good work, keep it up;" and Group III received specific comments that the teacher thought would be helpful. The subsequent criteria test, an objective test, showed that the highest scores were achieved by pupils who received specific comments, the next highest by those who received general comments, and the lowest scores by those who received no comments.

In an investigation with 233 seventh-grade students, Bridgeman (1974) found that "success" feedback was followed by significantly higher

performance than was "failure" feedback in students who had performed equally well on an earlier test. Scores from a test administered two days earlier were randomly assigned to the students immediately before taking the test. The results suggest that the student's perception of his prior success or failure influences his subsequent performance.

It is more effective to delay feedback from a test if one wants the greatest retention of learning. Brackbill and Kappy (1962) found that immediate reinforcement, as received on programmed materials or on a computer test, is not as effective as delayed feedback. After a test, it is better to wait at least 24 hours to go over the items. When feedback from a test is delayed by several days, learners perform better on a retest than when feedback is provided immediately or after 24 hours. Presumably, wrong answers interfere with learning and recalling the correct answer provided by the immediate feedback (Surber and Anderson, 1975).

Extensions and Exceptions

In reinforcement studies of lower class black males aged 9-11 years, Terrill, Terrill and Taylor (1980, 1981) found tangible reinforcement was more effective than other types of reinforcement and no reward. With mildly mentally retarded black males both the candy and culturally relevant praise such as "nice job, little brother" significantly increase their test scores.

Institutionalized educable mentally retarded students who received verbal praise and reinforcement by positive examiners scored significantly higher on the WISC-R Performance (Saigh, 1981).

Eighth grade social studies students who received knowledge of the correct response within a short interval after the response was emitted were compared with subjects who did not have knowledge of the correct response. Oliver and Taylor (1970) found higher retention scores for slow students who received feedback, but no significant difference for accelerated students.

REFERENCES

Anastasi, A., and Cordova, F.A. Some Effects of Bilingualism upon the Intelligence Test Performance of Puerto Rican Children in New York City. Journal of Educational Psychology (1953) 44:1-19.

Barringer, C., and Gholson, B. Effects of Type and Combination of Feedback upon Conceptual Learning by Children: Implications for Research in Academic Learning. Review of Educational Research (1979) 49:459-478.

Bauer, D. The Effects of Instructions, Anxiety and Locus of Control on Intelligence Test Scores. Measurement and Evaluation in Guidance (1975) 8:12-19.

Bligh, H. F. Trends in the Measurement of Educational Achievement. Review of Educational Research (1965) 35:34-49.

Brackbill, Y., and Kappy, M. S. Delay of Reinforcement and Retention. Journal of Comparative and Physiological Psychology (1962) 55:14-18.

Bridgeman, B. Effects of Test Score Feedback on Immediately Subsequent Test Performance. Journal of Educational Psychology (1974) 66:62-66.

Burr, W. L. Empirical Relationships Among Modes of Testing, Modes of Instruction and Reading Levels: In Sixth Grade Social Studies. Journal of Experimental Education (1963) 31:433-435.

Bushway, A., and Nash, W. R. School Cheating Behavior. Review of Educational Research (1977) 147:623-632.

Carrier, C. A., and Titus, A. Effects of Notetaking Pretraining and Text Mode Expectations on Learning from Lectures. American Educational Research Journal (1981) 18:385-397.

Clawson, T. W., Firment, C. K., and Trower, T. L. Test Anxiety: Another Origin for Racial Bias in Standardized Testing. Measurement and Evaluation in Guidance (1981) 13:210-215.

Cotler, S., and Palmer, R. The Effects of Text Anxiety, Sex of Subject, and Type of Verbal Reinforcement on Maze Performance of Elementary School Children. Journal of Personality (1970) 38:216-234.

Culler, R. E., and Holahan, C. J. Text Anxiety and Academic Performance: The Effects of Study-Related Behaviors. Journal of Educational Psychology (1980) 72:16-20.

Curtis, H. A., and Kropp, R. P. A Comparison of Scores Obtained by Administering a Test Normally and Visually. Journal of Experimental Education (1961) 29:249-260.

Dreisbach, M., and Keogh, B. K. Testwiseness as a Factor in Readiness Test Performance of Young Mexican-American Children. Journal of Educational Psychology (1982) 74:224-229.

Fischer, C. T. Levels of Cheating under Conditions of Informative Appeal to Honest, Public Affirmation of Value and Threats of Punishment. Journal of Educational Research (1970) 64:12-16.

French, J. W. Effect of Anxiety on Verbal and Mathematical Examination Scores. Educational and Psychological Measurement (1962) 22:553-564.

Gorden, L. V., and Durea, M. A. The Effect of Discouragement on the Revised Stanford-Binet Scale. Journal of Genetic Psychology (1948) 73:201-207.

Green, K. Item-Response Changes on Multiple-Choice Tests as a Function of Test Anxiety. Journal of Experimental Education (1981) 49:225-228.

Hughes, D. C. An Experimental Investigation of the Effects of Pupil Responding and Teacher Reacting on Pupil Achievement. American Educational Research Journal (1973) 10:21-37.

Ingle, B., and DeAmico, G. Effect of Physical Conditions of the Test Room on Standardized Achievement Test Scores. Journal of Educational Measurement (1969) 6:237-240.

Jensen, A. R. The Effect of Race of Examiner on the Mental Test Scores of White and Black Pupils. Journal of Educational Measurement (1974) 11:1-14.

Johnson, R. E., and Klores, M. S. Attitudes Toward Cheating as a Function of Classroom Dissatisfaction and Peer Norms. Journal of Educational Research (1968) 62:60-64.

Kalechstein, P., Kalechstein, M. and Docter, R. The Effects of Instruction on Test-Taking Skills in Second Grade Black Children. Measurement and Evaluation in Guidance (1981) 13: 198-202.

Kalish, R.A. An Experimental Evaluation of the Open Book Examination. Journal of Educational Psychology (1958) 49: 200-204.

Kirkland, K., and Hollandsworth, J.G., Jr. Test Anxiety, Study Skills, and Academic Performance. Journal of College Student Personnel (1979) 20: 431-436.

Kirkland, M.C. The Effects of Tests on Students and Schools. Review of Educational Research. (1971) 41: 303-351.

Marsh, R. Should We Discontinue Classroom Tests? An Experimental Study. High School Journal (1980) 63: 288-292.

McCarthy, D. A Study of the Reliability of the Goodenough Drawing Test of Intelligence. Journal of Psychology (1944) 18:201-216.

McQueen, R. Examination Deception as a Function of Residual Background and Immediate Stimulus Factors. Journal of Personality (1957) 25:643-650.

Mims, M., and Gholson, B. Effects of Type and Amount of Feedback upon Hypothesis Sampling Among 7-8-Year Old Children. Journal of Experimental Child Psychology (1977) 24:358-371.

Mishra, S. P. The Influence of Examiners' Ethnic Attributes on Intelligence Test Scores. Psychology in the Schools (1980) 17:117-122.

Naughton, J. A Modest Experiment on Test Motivation. Personnel and Guidance Journal (1968) 46:606.

Nottelman, E. D., and Hill, K. T. Test Anxiety and Off-Task Behavior in Evaluative Situations. Child Development (1977) 48:225-231.

Oakland, T. The Effects of Testwiseness Materials on Standardized Test Performance of PreSchool Disadvantaged Children. Journal of School Psychology (1972)10:355-360.

Oliver, J. K., and Taylor, B. L. The Effects of Teaching Social Studies on Immediate Retention. Journal of Educational Research (1970) 63:198-200.

Osler, S. F. Intellectual Performance as a Function of Two Types of Psychological Stress. Journal of Experimental Psychology (1954) 47:115-12[illegible].

Page, E. B. Teacher Comments and Student Performance: A Seventy-Four Classroom Experiment in School Motivation. Journal of Educational Psychology (1958) 49:173-181.

Ratliff, R. G. Two-Choice Discrimination Learning in Children as a Function of Punishment Modality and Reinforcement Combination. Journal of Experimental Child Psychology (1972) 14:365-371.

Reichenberg-Hackett, W. Changes in Goodenough Drawings After a Gratifying Experience. American Journal of Orthopsychiatry (1953) 23:501-517.

Rindler, S. E. The Effects of Skipping Over More Difficult Items on Time-Limited Tests: Implications for Test Validity. Educational and Psychological Measurement (1980) 40:989-998.

Rogosin, H. What About "Cheating" on Examinations and Honesty? School and Society (951) 74:402-403.

Saigh, P. A. The Effects of Positive Examiner Verbal Comments on the Total WISC-R Performance of Institutionalized EMR Students. Journal of School Psychology (1981) 19:86-91.

Schwarz, J. C. A New Procedure for Administering Objective Tests to Large Classes. Journal of Educational Measurement (1967) 4:167-168.

Shannon, A. J. Effects of Methods of Standardized Reading Achievement Test Administration on Attitudes toward Reading. Journal of Reading (1980) 23:684-686.

Shirk, E., and Hoffman, R. W. The Academic Setting of the Dishonest Student. Improving College and University Teaching (1961) 9:130-134.

Slakter, M. J., Koehler, R. A., and Hampton, S. H. Grade Level, Sex and Related Aspects of Testwiseness. Journal of Educational Measurement (1970) 7:119-122.

Spence, J. T. The Distraction Effects of Material Reinforcers in the Discrimination Learning of Lower and Middle Class Children. Child Development (1970) 41:103-111.

Steininger, M., Johnson, R. E., and Kirts, D. K. Cheating on College Examinations as a Function of Situationally Aroused Anxiety and Hostility. Journal of Educational Psychology (1964) 55:317-324.

Super, D. E., Braasch, W. F., Jr., and Shay, J. B. The Effects of Distractions on Test Results. Journal of Educational Psychology (1947) 38:373-377.

Surber, J. R., and Anderson, R. C. Delay-Retention Effect in Natural Classroom Settings. Journal of Educational Psychology (1975) 67:170-173.

Swanson, E. N., and Deblassie, R. R. Interpreter and Spanish Administration Effects on the WISC Performance of Mexican-American Children. Journal of School Psychology (1979) 17:231-236.

Szafran, R. F. Question-Pool Study Guides: Effects on Test Anxiety and Learning Retention. Teaching Sociology (1981) 9:31-43.

Terrill, F., Terrill, S.L., and Taylor, J. Effects of Race of Examiner and Type of Reinforcement on the Intelligence Test Performance of Lower-Class Black. Psychology in the Schools (1980) 17: 270-272.

Traxler, A. E., and Hilkert, R. N. Effect of Type of Desk on Results of Machine-Scored Tests. School and Society (1942) 56:277-296.

Trentham, L. L. The Effect of Distractions on Sixth-Grade Students in a Testing Situation. Journal of Educational Measurement (1975) 12:13-18.

Trentham, L. L. Anxiety and Instruction Effects on Sixth-Grade Students in a Testing Situation. Psychology in the Schools (1979) 16:439-443.

Uhlig, G. E., and Howes, B. Attitude Toward Cheating and Opportunistic Behavior. Journal of Educational Research (1967) 60:411-412.

Vitro, F. T., and Schoer, L. A. The Effects of Probability of Test Success, Test Importance, and Risk of Detection on the Incidence of Cheating. Journal of School Psychology (1972) 10:269-277.

Walker, R. E., Sannito, T. C., and Firetto, A. C. The Effect of Subjectively Reported Anxiety on Intelligence Test Performance. Psychology in the Schools (1970) 7:241-243.

Wine, J. Test Anxiety and Direction of Attention. Psychological Bulletin (1971) 76:92-104.

Wrightsman, L. S., Jr. The Effects of Anxiety, Achievement Motivation, and Task Importance upon Performance on an Intelligence Test. Journal of Educational Psychology (1962) 53:150-156.

Yamamoto, K., and Dizney, H. F. Effects of Three Sets of Test Instructions on Scores on an Intelligence Scale. Educational and Psychological Measurement (1965) 25:87-94.

FORMATIVE OBSERVATION INSTRUMENT

6.0 TESTING: STUDENT PREPARATION, ADMINISTRATION, FEEDBACK

Florida Performance Measurement System
Coalition for the Development of a Performance Evaluation System
Office of Teacher Education, Certification and Inservice Staff Development
Tallahassee, Florida

IDENTIFICATION AND FORMAT INFORMATION

Teacher's Name:_____ Date: _____

School: _____ Department/Team: _____

Observer's Name:_____ Observation Number:_____

Starting Time:_____ Length of Observation:_____

Brief Statement of Lesson Purpose or Content:_____

_____

_____

This instrument is designed to record effective and ineffective indicators of teacher behavior in the domain of Testing: Student Preparation, Administration, Feedback. The instrument is divided into three categories: Preparation For Testing; Test Administration; and Formative Feedback.

Directions:

1. Place a mark in the appropriate box when a relevant behavior is observed. (Effective items are on the left of the instrument and ineffective items are on the right.)
2. Mark an item each time it is observed.
3. Sum frequencies by indicators, record subtotals for each category and record an effective and ineffective total for the categories in the appropriate spaces provided below.

Note: Items on the Observation Instrument are observed at varying times. They are as follows:

6.1 Preparation for Testing. Observed one to three days before testing.

6.2 Test Administration. Observed during the period students take an examination.

6.3 Formative Feedback. Observed two or more days after an examination.

Data Summary

CATEGORY	EFFECTIVE	INEFFECTIVE
6.1 Preparation for Testing		
6.2 Test Administration		
6.3 Formative Feedback		
TOTAL	185	

DOMAIN 6.0 OBSERVATION INSTRUMENT

CATEGORY	EFFECTIVE INDICATOR	FREQUENCY	FREQUENCY	INEFFECTIVE INDICATOR
6.1 PREPARATION FOR TESTING (Observed prior to testing.)	States test purpose(s)			
	Explains use of test results			Relates test purpose to grading
	Explains importance of results			Threatens reprisals for failure
	States positive expectations			Makes negative comments about student ability
	States content to be tested			Uses vague/general terms to describe test content or review
	Reviews test content			
	Cites study references or provides study guide			Emphasizes test difficulty
	Describes item type(s)			
	Gives test-taking directions			Makes indefinite statements about test format/directions
	Explains test-taking strategies			
	Provides test-taking practice			
6.2 TEST ADMINISTRATION (Observed during testing period.)	Arranges chairs to provide space among students			Allows close proximity
	Controls temperature and ventilation			Disregards room temperature extremes
	Provides lighting			Disregards shadows, reflections or other poor lighting
	Clarifies student questions			Refuses to help or respond to test related questions
	Offers assistance to individuals with questions			
	Monitors students while taking exam			Leaves the room during testing
	Controls noise and interruptions			Occupied with other tasks/loses awareness of student behavior
6.3 FORMATIVE FEEDBACK (Observed when test results are returned.)	Reviews test items/gives reason for answers			Uses tokens as feedback
	Listens to alternative responses			
	Clarifies misconceptions			
	Makes positive comments about student ideas			

Section II, Appendix C

DOMAIN

7.0 ONE-ON-ONE SESSION

7.0 ONE-ON-ONE SESSION

Definition: This domain = behavior of instructional personnel such as school counselor, social worker, psychologist, peer teacher, and others (referred to hereafter as helper) who provide help to a learner in a one-on-one setting.

This domain consists of the following named concepts:

7.1 Efficient Use of Time

7.2 Development of Rapport

7.3 Communication Skills

7.4 Interpersonal Skills

7.5 Problem Formulation

7.6 Development of Plans

Note: This Domain is tentative at this time. It must be tested in the one-on-one setting and appropriately revised. For research findings related to the concepts in this document, refer to Domains 3, 5, and 8.

7.1 Concept: Efficient Use of Time

Definition: Efficient use of time = helper behavior that maximizes the use of allocated time for learning.

Indicators

7.1.1 Punctuality

Definition: Punctuality = helper begins session promptly.

7.1.2 Controlled Interruptions

Definition: Controlled interruptions = helper provides procedures to be followed by anyone who must contact either conferee.

7.1.3 Housekeeping

Definition: Housekeeping = helper has materials prepared, procedures worked out, and everything in order.

7.2 Concept: Develops Rapport

Definition: Develops rapport = helper behavior that initiates and sustains a relationship of mutual confidence and trust with a learner.

Indicators

7.2.1 Arranges Social Space

Definition: Arranges social space = helper arranges seats so as to reduce the social distance between helper and learner.

7.2.2 Promptness

Definition: Promptness = helper begins the meeting promptly.

7.2.3 Explains Purpose of Encounter

Definition: Explains purpose of encounter = helper explains that the talk is to help the learner with problems of teaching and/or learning.

7.2.4 Elicits Learner's Concerns

Definition: Elicits learner's concerns = helper asks questions that provide opportunity for the learner to freely express feelings, interest, or difficulties.

7.3 Concept: Communication Skills

Definition: Communication skills = verbal and nonverbal behavior that evokes and expresses information and personal relationships

Indicators

7.3.1 Connected Discourse

Definition: Connected discourse = thematically connected discourse that leads to at least one point.

7.3.2 Scrambled Discourse

Definition: Scrambled discourse = discontinuous or garbled verbal behavior in which ideas are loosely associated.

7.3.3 Vagueness Words

Definition: Vagueness words = words of everyday speech which persons often use to describe, to present information, or to answer questions and for which the denotations are indeterminate (e.g., something, a little, some, much, few, things, you see, perhaps, actually).

7.3.4 Question Overload

Definition: Question overload = helper verbal performance in which long involved questions or multiple questions are asked.

7.3.5 Single Questions

Definition: Single questions = helper asks direct questions one at a time without rephrasing or giving additional information.

7.3.6 Listens

Definition: Listens = helper gives undivided attention to what the learner says, responding in a way that indicates to the learner that s/he was heard.

7.3.7 Provides Pause Time

Definition: Provides pause time = helper paces interaction to allow time for learner to deliberate.

7.3.8 Pacing

Definition: Pacing = helper focusses on a discussion topic for an extended period, long enough to explore it in some depth.

7.3.9 Marker Expressions for Emphasis

Definition: Marker expressions for emphasis = helper use of words (e.g., note this, this is the most important point) to indicate that to which the learner should give special attention.

7.3.10 Marker Techniques for Emphasis

Definition: Marker techniques for emphasis = helper performance that uses underlining, colors, cartoons, etc., in presenting information.

7.3.11 Repetition

Definition: Repetition = helper performance in which the main points are restated at spaced intervals.

7.3.12 Helper's Speech

Definition: Helper's speech = voice characteristics that make up the auditory stimuli as distinguished from the content or message of the discourse such as: loud, noisy or grating voice; shrill, piercing, highly pitched voice; monotone--fails to vary the intensity, rate, and volumne of speech; and speaks too softly, almost inaudibly.

7.3.13 Body Language

Definition: Body language = helper's facial or other body behavior that expresses interest, excitement, joy, and positive personal relations; or boredom, sadness, dissatisfaction, or negative personal relations; or else, no clear message at all, such as: smiles, deadpan expression, frowns, posture, movement and eye contact.

7.4 Concept: Interpersonal Skills

Definition: Interpersonal skills = verbal behavior that enables the helper to convey information and to control the quality and direction of interaction.

Indicators

7.4.1 Minimizes Evaluative Expression

Definition: Minimizes evaluative expression = helper uses descriptive language, relatively free of evaluative judgments, in expressing inferences, giving or calling for information and summarizing.

7.4.2 Gives Corrective Feedback

Definition: Gives corrective feedback = helper makes statements or asks questions that provide constructive suggestions for changes in either comprehension or behavior.

7.4.3. Gives Specific Reinforcement

Definition: Gives specific reinforcement = helper behavior that tends to increase the chances that the learner will be both pleased and likely to repeat the behavior.

7.4.4 Exercises Flexibility

Definition: Exercises flexibility = helper sensitivity to a range of ways in which behavior can be interpreted.

7.4.5 Maintains Problem Focus

Definition: Maintains problem focus = helper behavior that guides the problem-solving process by eliciting relevant information or by giving such information.

7.5 Concept: Problem Formulation

Definition: Problem formulation = helper behavior that involves the teacher in identifying changes in behavior which would eliminate differences between the present and desired behavior.

Indicators

7.5.1 Assess Data Needs

Definition: Assess data needs = helper behavior that involves the learner in determining what data would provide useful information about learner behavior.

7.5.2 Examines Product Data

Definition: Examines product data = helper behavior that involves the learner in a cooperative examination of a work sample of the learner's performance in order to identify strengths and weaknesses.

7.5.3 Examines Test Data

Definition: Examines tests data = helper behavior that involves the learner in a cooperative examination of teacher made or standardized tests in order to assess level of learner ability, student achievement level, and/or test characteristics.

7.5.4 Examines Observation Data

Definition: Examines observation data = helper behavior that involves the learner in a cooperative examination of the data coded on an observation instrument in order to identify effective and ineffective behavior.

7.5.5 Elicits Learner's Concerns about Data

Definition: Elicits learner's concerns about data = helper questions or statements that induce the learner to talk about the strengths and weaknesses of his or her behavior (based on data).

7.5.6 Seeks Clarification of Concerns

Definition: Seeks clarification of concerns = helper tests his or her understanding of concerns by repeating learner's statement(s) or by giving interpretations of what learner said.

7.5.7 Attends to Unnoticed Data

Definition: Attends to unnoticed data = helper questions or statements that call attention to behavior, especially ineffective behavior.

7.5.8 Elicits Summary of Data Analysis

Definition: Elicits summary of data analysis = helper questions or statements that induce the learner to summarize the results of the data analysis.

7.5.9 Restates Summary

Definition: Restates summary = helper states summary of analysis or asks questions to establish agreement regarding the behavior to be targeted for change or developed outright.

7.6 Concept: Development of Plans

Definition: Development of plans = helper behavior that engages the learner in the formulation of plans to improve performance of the targeted behavior.

Indicators

7.6.1 Elicits Analysis of Targeted Behavior

Definition: Elicits analysis of targeted behavior = helper questions or statements that invite the learner to express his or her understanding of the behavior.

7.6.2 Checks for Agreement

Definition: Checks for agreement = helper states his or her interpretation of the behavior or restates what the learner said in order to establish agreement between helper and learner.

7.6.3 Initiates Planning

Definition: Initiates planning = helper makes statements or asks questions that stimulate the learner to join in suggesting ways to change behavior.

7.6.4 Summarizes Plans

Definition: Summarizes plans = helper asks questions or makes statements to induce learner to restate plans.

7.6.5 Confirms Plans

Definition: Confirms plans = helper makes statements or asks questions to test agreement on plans, including materials and activities, time for next observation and interim conference.

EXAMPLES FOR DOMAIN 7.0

Directions: Here are a few examples to exemplify some of the behaviors that are defined in this domain. As you read through the documents list the indicator numbers in the blanks. The correct answers are given on the last page.

_____ 1. The school psychologist set up a meeting with Tommy's father to discuss the results of a staffing. She greeted Mr. Johnson and invited him to sit in the chair she had placed at a right angle next to her desk.

_____ 2. The Learning Disabilities teacher has a group of appropriate books ready for Gina to make a selection for her practice reading. He has also prepared flashcards and dittoed exercises for drill. They will use a reading accelerator which he has checked to see if it is operating correctly. All materials are ready and organized in the sequence he has planned for Gina's lesson.

_____ 3. When John came to the conference room, the librarian was ready to begin. She had materials organized, and began the session promptly by greeting John and inviting him to sit next to her at the conference table. She then began the discussion by stating, Mr. McQuire told me that you would like to explore some possible careers in woodworking. Can we begin by you telling me more about your interest in this?

_____ 4. The social worker instructed the secretary to hold telephone calls, visitors, and other types of interruptions until the end of the scheduled session. He also posted a "do not disturb" sign on his door.

_____ 5. Helper: Tell me about your concerns regarding your students.
Ms. Bee: Well, things get out of control pretty quickly. I can't figure out what sets them off but there must be some ringleaders who carry the class along. If I just knew how to prevent the problem from starting...
Helper: Uh hum. You feel the class control has moved from you to a few students. Is that right? Tell me what you mean by things being out of control?

_____ 6. Example (Neg): Teacher: I just don't seem to do anything right. No matter how hard I try I am not able to follow the lesson plans I write for my class and the principal criticizes the way I conduct the lesson. I'm afraid I will lose my job.
Peer teacher: (Pause) Well, let's see what we planned to discuss this week. Did you bring the data you collected on the number of rule infractions you handled successfully?

_____ 7. Supervisor: The observation tallies for praise indicates that you tend to use general praise with the students. While this does provide reinforcement, studies of effective teachers have shown that specific praise is more effective, particularly because it cues the students into exactly what it is that they did that was good and avoids misinterpretations by them.

_____ 8. Peer teacher: These lesson plans that you wrote for presentation of value knowledge are excellent. . .you specified what criteria you want to emphasize and listed clear examples and nonexamples. Did you find this helped you in the class?

_____ 9. Supervisor: The observation sheet shows that the lesson frequently gets sidetracked by student questions which are irrelevant. What we need to do is come up with a way you can be comfortable with of getting the discussion back on track without making the students feel discouraged from participating.

_____ 10. Peer teacher: You said that you continually have difficulty getting the class to finish the classwork you assign. What are the ways we could check first to determine if the problem lies in the way you handle the class or the clarity of the explanation or the amount of work you expect them to do?

_____ 11. Counselor: I have collected the standardized scores of your class so that we can refer to any student's latest measure of performance on the Stanford Achievement tests in mathematics. It may help us as we look at your mathematics problems and identify any areas you think are difficult for most of your students.

_____ 12. Helper: According to what you said, an effective desist should focus on the correct person, stop the misbehavior from spreading, not interrupt the lesson flow too severely and not be harsh in nature. Did I leave anything out?

_____ 13. Learning Disabilities teacher: To make sure we both understand what the improvement plan will be, let's briefly review the plan and write it down, indicating what each of our responsibilities are. So, for tomorrow what's the first step?

_____ 14. Helper: Are we agreed that we plan to concentrate on getting control of the classroom again even if some measures seem in conflict with your belief in student-centered operation? You said you would try to explain the need to change classroom procedures that are causing disruptions and elicit the students' help in developing a few rules that all of you could accept. You plan to closely monitor the rules and provide punishment for any infraction. Is that correct? (Pause) We plan to meet briefly at 3:15 p.m. to discuss any difficulties.

CORRECT ANSWERS

1. 7.2.1
2. 7.1.3
3. 7.1.1
4. 7.1.2
5. 7.2.4
6. 7.3.6 N
7. 7.4.1
8. 7.4.3
9. 7.4.2
10. 7.5.1
11. 7.5.3
12. 7.6.2
13. 7.6.4
14. 7.6.5

Section II, Appendix D

SPECIALIZED DOMAIN
(Domain Pertaining to a Speciality)

8.0 COMMUNICATION OF OBSERVATION DATA

8.0 COMMUNICATION OF OBSERVATION DATA

Definition: This domain = behavior of principal, supervisor, or peer teacher (referred to hereafter as consultant) who provides help to a teacher whose need is decided on the basis of data derived from observation of classroom performance.

This domain consists of the following named concepts:

8.1 Development of Rapport

8.2 Interpersonal Skills

8.3 Problem Formulation

8.4 Development of Plans

Overview

Several terms are used to designate the individual who gives assistance to a teacher on instructional problems. Among these terms are "conferrer," "resource teacher," "consultant," "instructional consultant," and "guidance consultant" (see Good's Dictionary of Education 1959). "Consultant" is the term typically used in the research literature, and it is used in this domain where the research is reviewed and where the principal, supervisor, or peer teacher functions in a helping capacity. In a pedagogical sense, consultant is a person who serves as an advisor or helper in reference to problems of planning, instruction, evaluation, and student conduct.

There are at least two types of consulting situations, as follows:

Type 1. In this type, a teacher who has a problem with instruction or student conduct works with a consultant to solve the problem. Data are collected about the class or a particular student, depending upon the problem, and in terms of these data and what the consultant knows about psychology or pedagogical research, they together work out plans to be implemented as a solution. In this case, the primary attention is focussed on data about the class or student. Once the problem is defined attention is given to the question of how to cope with it.

Type 2. This type is one in which the classroom performance of the teacher is the primary concern. Data are collected on the teacher's performance by observation and the teacher together with the consultant analyze the data to determine the performance that needs improvement. The focus here is on data about the teacher's performance rather than on data about the student or class. Nevertheless, the students are the ultimate beneficiaries of whatever improvements are made in the teacher's performance.

Naturally, the teacher will join in consultation with more concern about classroom difficulties than with his or her own behavior. This preoccupation of the teacher with classroom problems places restraint upon the consultant, for these pressing concerns cannot be brushed aside without alienating the teacher. Attention must be given to the classroom, and in doing so the classroom difficulty can be associated with the observed performance of the teacher. In other words, the consulting situation may begin with Type 1 and move to Type 2.

Consulting is a process whose purpose is to help other professionals engage in more effective behavior. There is perhaps no function that principals, supervisors, or peers can perform that can more directly affect the learning environment of the school than consulting with teachers who need help. Yet to be successful in the helping relationships requires that the consultant be sensitive to the teacher's concerns, attitudes, perspectives, and problems; that the consultant understand the consulting process and possess the skills of interpersonal communication; and that the consultant have the pedagogical knowledge and skills to provide the specific help the teacher needs.

Consultation in this document is distinguished from counseling by the type of service provided by the consultant (principal, peer teacher, others) to the teacher, otherwise called the consultee. The focus of the consultation activities is to provide assistance to the teacher by sharing information. In contrast, counseling would deal more with the teacher's emotions and affective adjustments and social problems. In the event assistance is needed in this area, the consultant should help refer the teacher to the appropriate person/agency (unless the consultant is appropriately trained to render counseling services in which case the liason becomes therapy and not consultation).

The research indicates that effective consultation requires skilled consultants. What are the skills the principals, peer teachers and others should have in order to be effective consultants?

It is the consultant's responsibility to guide the teacher through consideration of the data, problem formulation and development of plans to solve the problem. The consultant does this largely through skills of communication that enhance interpersonal relations. Successful communication will usually focus on the data and the problem, rather than on the teacher <u>per se</u>. For example, "Tell me what areas of Presentation of Subject Matter seem to be effective and what areas are ineffective according to the data on this observation sheet" is a request that focusses the conference on a specific analysis of data and is preliminary to planning. It can be delivered in other ways, for example, asking the teacher questions to cue him/her to the areas of concern, e.g., "From the observation data on Presentation of Subject Matter have you noticed any increase in tallies?"

This Domain is designed to provide knowledge about consulting and to awaken and heighten the sensitivity of principals, supervisors, and peers to the feelings and concerns of the teachers with whom they work.

Because consulting is seen as a process of helping others engage in more rewarding performance, basic differences arise over the question of how to approach and react to those who need help. These differences are essentially the same as those which swirl about the issue of didactic and heuristic teaching. Is it better to use a direct or a collaborative approach? Some say, be direct; others say collaborative. The research on this issue is also reviewed as well as research on the question of whether or not consulting is effective in changing teacher performance.

Concept: Development of Rapport

Definition: Development of rapport = Consultant behavior that initiates and sustains a relationship of mutual confidence and trust with a teacher.

Indicators

8.1.1 Arranges Social Space

Definition: Arranges social space = consultant arranges seats so as to reduce the social distance between consultant and teacher.

Example (Pos)*: When Ms. Jones arrives at her peer teacher's classroom to discuss the observation results, Ms. Holbrook is seated at the desk. She asks Ms. Jones to sit in the chair at the end of the desk with the observation results positioned between the two teachers so that both can see them.

Example (Neg): The principal, Ms. Brown, asks Ms. Thompson to come to her office to discuss the observation results. When Ms. Thompson enters the office, Ms. Brown is seated behind her desk with the results in front of her and an empty chair is at the other side of the desk. The principal motions for Ms. Thompson to sit.

8.1.2 Promptness

Definition: Promptness = consultant is on time for appointments with the teacher.

Example (Pos): Mr. Profetto's supervisor, Ms. Chartier, set up an appointment to discuss the observation results at 2:30. He arrived at Ms. Chartier's classroom promptly at 2:30, equipped with all the necessary papers and materials.

Example (Neg): Ms. Casey has an appointment to confer with her peer teacher, Ms. Cuny at 3:15. When Ms. Casey arrives at Ms. Cuny's classroom for the meeting, Ms. Cuny is busy putting away materials from the school day and chatting with a few students who hang around after class. Ms. Cuny finally acknowledges the presence of Ms. Casey and asks her to have a seat. Ms. Cuny asks the students to leave, then starts gathering materials for the conference with Ms. Casey. By 3:45 she is ready to begin. She apologizes for the delay.

*"Pos" indicates an example that has positive effects; "Neg," an example that has negative effects.

8.1.3 Controls Interruptions

Definition: Controls interruptions = consultant provides procedures to be followed by anyone who must contact either the consultant or consultee.

Example (Pos): Ms. MacDowell, the principal, set up an appointment with Ms. Anderson to discuss the observation results at 3:30. When Ms. Anderson arrived at 3:30, Ms. MacDowell was waiting with the summary sheet available, and she asked the secretary to hold her calls until the meeting was through.

Example (Neg): The conference with Mr. Williams was interrupted by the secretary presenting letters for him to sign and messages from parents and teachers. At one point, he abruptly left the office during the conference and when he returned Ms. Baker had to review the previous statements to help him get back to the purpose of their discussion.

8.1.4 Explains Conference Purpose

Definition: Explains conference purpose = consultant explains that the conference is to help the teacher with problems of teaching and classroom management.

Example (Pos): Ms. Ramos, the peer teacher, says to Ms. Denburg: The purpose of this conference is to decide how I can best help you to improve your teaching skills. We're going to look over the results of the observations. While we're reviewing them, we'll be able to see which skills are strong points, and which ones we can work on to improve. We'll work together to decide what areas might be causing difficulty and to develop a plan to cope with them.

Example (Neg): Ms. Lefebvae's supervisor says to her: Okay, let's see what sort of problems you've been having in the classroom. (She then proceeds to point out those areas of the observation results which show a need for improvement.)

8.1.5 Minimizes Personal Information

Definition: Minimizes personal information = consultant minimizes questions that solicit personal information from the teacher.

Example (Pos): Consultant: Do you want to talk about your concerns regarding student conduct in your classroom?
Ms. Bee: Well, things get out of control pretty quickly. I can't figure out what sets them off but there must be some ringleaders who carry the class along. If I just knew how to prevent the problem from starting...

Consultant: Uh hum. You feel the class control has moved from you to a few students. Is that right? (Pause) What do you mean by things being out of control?
Ms. Bee: Whenever I try to involve the class in a discussion, they shout out the answers to questions I ask. And it ends up in a shouting match. I am really afraid to stop lecturing and try to have a discussion.
Consultant: One way to gain control of this situation would be to take time to establish some rules before the question asking. You know, discuss the problem with the class first.

Example (Neg): Consultant: We're going to review the results of the observations and see how well you're doing in the classroom. I noticed you were upset in class today. Are you usually upset when your own children are unruly?

8.1.6 Body Language

Definition: Body language = consultant's facial or other body behavior that expresses interest, excitement, joy, and positive personal relations; or boredom, sadness, dissatisfaction, or negative personal relations; or else, no clear message at all, such as: smiles, deadpan expression, frowns, posture, movement, and eye contact.

Example (Neg): The consultant gives an incongruent message to the teacher by professing interest in the teacher's work and, at the same time, keeps glancing at the clock or reading messages on his/her desk.

Principle

If social arrangements are attended to and requests for personal information is held to a minimum, rapport is likely to be influenced. We found no empirical evidence to support indicators 8.1.2, 8.1.3, 8.1.4.

Supportive Evidence

Broekmann and Moller (1973) investigated the preference for four different seating arrangements, as well as the preferred distance between chairs in each position, in each of the following situations: 1) counseling, 2) formal (job application), 3) social-general (visit with friend of same sex), and 4) home (conversation with a parent). The position of chairs opposite each other with a table intervening was preferred significantly more often in counseling, formal, and social-specific situations. The position of chair side by side at a 45 degree angle was preferred most often in home situation. The middle distance (1.5 meters) was preferred most often in all situations.

Haase and Di Mattia (1976) found the preferred arrangement within each group (counselors, administrators, and clients) was arrangement 3, in which the participants interract over the corner of the desk. However, there were differences in magnitude of preference. The greatest differences between groups were found for arrangement 4 (table intervening). Of the three groups, counselors preferred arrangement 4 the least, administrators the most, and clients fell between the other two groups. In addition, arrangement 1 (two chairs placed side by side at a 45 degree angle, no desk intervening) was preferred by the clients the least of all 4 arrangements, while counselors preferred it the most of all 4 arrangements.

Greene (1977) studied the effects of a counselor's verbal feedback and proximity to a client on eliciting behavioral compliance. At the personal distance (2 feet), positive feedback was more effective than neutral feedback; while at the social distance (5 feet), neutral feedback was more effective than acceptance in evoking compliance. In these two successful conditions, clients tended to feel less distrustful. The author speculates that these two conditions "can be characterized as one in which the counselor's communication of attitude was consistent across verbal and nonverbal channels," and that perceived inconsistencies make the client suspicious and reduce the counselor's influence. Follow-up questionnaires also indicated that clients who felt too much personal information had been asked for tended to be less compliant.

8.2 Concept: Interpersonal Skills

Definition: Interpersonal skills = verbal behavior that enables the consultant to convey information and to control the quality and direction of interaction.

Indicators

8.2.1 Minimizes Evaluative Expression

Definition: Minimizes evaluative expression = consultant uses descriptive language, relatively free of evaluative judgments, in expressing inferences, giving or calling for information, and summarizing.

Example (Pos.): Consultant: The tallies for praise indicate that you tend to use general praise with the students. While this does provide reinforcement, research has shown that specific praise is more effective, particularly because it cues the students into exactly what it is that they did that was good and avoids misinterpretation by them.

Example (Neg.): Consultant: According to the observation data, you have poor classroom management skills. This is obviously an important concern and will need to be improved.

8.2.2 Gives Corrective Feedback

Definition: Gives corrective feedback = consultant makes statements or asks questions that provide constructive suggestions for changes in either comprehension or performance of an indicator(s).

Example (Pos.): Consultant: Since we've noticed that there are tallies in both the effective and ineffective columns of this indicator, maybe we should first try to determine in what circumstances each type occurs. One way we could gather the data is for you to try keeping a record of what is happening when you find yourself having difficulty in this area....the type of activity going on, which students, time of day....that sort of thing.

Example (Neg.): Consultant: You say that the major difficulty as far as behavior control is not how to deal with major disruptions but how to control minor disturbances before they spread and become problems. Your discipline seems to be improving. There haven't been any referrals to the principal this week, and I haven't seen any kids standing in the hallway outside your door. That's just fine. Keep doing whatever you are doing.

8.2.3 Gives Specific Reinforcement

Definition: Gives specific reinforcement = consultant behavior that tends to increase the chances that the teacher will be both pleased and likely to repeat the behavior.

Example (Pos.): Consultant: You did an excellent job of keeping the momentum of the lesson flowing without letting minor incidents get out of hand and distract all the students. The way you circulated around the room while asking questions kept everyone's attention. It gave you an opportunity to stop misbehavior nonverbally like when you removed Joe's comic book, and touched those girls who were whispering.

Example (Neg.): Consultant: You are really doing a good job.

8.2.4 Exercises Flexibility

Definition: Exercises flexibility = consultant sensitivity to a range of ways in which an indicator can be interpreted or performed.

Example (Pos.): Consultant: We have discussed the fact that the observation data show that there seems to be time wasted on housekeeping details at the beginning of class that could be used more effectively for instruction. Maybe this could be improved by organizing your material differently. You have seen my method of using student group leaders to pass out materials which I have presorted and stacked according to student work groups. Some teachers use individual files for collecting and returning work. If you pass out work individually so that you can talk to each student, then you must provide a student activity or you will waste their time. Why don't you talk to a number of teachers who you respect and find out what sort of organization they use. That might give you some idea of a number of ways you could set things up....one of which you might be able to modify to suit your circumstances.

Example (Neg.): Consultant: I have told you that the only way you will have any discipline in your room is to always post a list of do's and don'ts on the front door. I'll give you a copy of my rules to post.

8.2.5 Maintains Problem Focus

Definition: Maintains problem focus = consultant focusses on a discussion topic for an extended period, long enough to explore it in some depth.

Example (Pos.): The consultant maintains control of the discussion by guiding the teacher through the problem-solving process. The content is controlled by the consultant verbalizations that call for a response from the teacher about a particular topic and the counselor does not shift content.

8.2.6 Listens

Definition: Listens = consultant gives undivided attention to what the teacher says, responding in a way that indicates to the teacher that s/he was heard.

Example (Pos.): Teacher: I know that I shouldn't encourage the students to raise irrelevant questions during a discussion by responding to them, but I don't want them to feel that I'm not interested in what they have to say.
Consultant: I think I understand. It is important for the students to feel you value their participation. Maybe you could respond with a comment like "That's an interesting question, Timmy, but right now we're discussing....But let's discuss that after this lesson is done." .... and then make a note to get back to it later. That way the student knows you're interested but the lesson doesn't get disrupted.

Example (Neg.): Teacher, I just don't seem to do anything right. No matter how hard I try I am not able to follow the lesson plans I write for my class and the principal criticizes the way I conduct the lesson. I'm afraid I will lose my job.
Consultant: (Pause) Well, let's see what we planned to discuss this week. Did you bring the data you collected on the number of rule infractions you handled successfully?

8.2.7 Provides Pause Time

Definition: Provides pause time = consultant paces interaction to allow time for teacher to deliberate.

Example (Pos.): Consultant: Why don't you take a few minutes to review the data and decide which areas you feel you are the most important to develop plans for. (Consultant waits for the teacher to look over the data and respond.)

Example (Neg.): Consultant: What area do you feel has given you the most difficulty? It seems to me that the most important concern is explanation of cause and effect relationships.

Principle

If the consultant maintains problem focus, is sensitive to a range of ways to cope with a problem, provides feedback and specific praise and is non-evaluative in discussion, then the likelihood of positive results from consultation is increased. We found no direct evidence in support of Indicators 8.2.6 and 8.2.7.

Supportive Evidence

Bergan and Tombari (1976) have identified consultation practices and skills from the literature (Caplan, 1970; Meyers, 1973, Bergan and Tombari, 1975) and conducted studies which support the effectiveness of such variables as consultant flexibility in the application of psychological principles, content relevance, interview focus, verbal processes, and message control.

In a study of flexibility, Bergan and Tombari (1976) defined flexibility as the use of a variety of psychological practices by the consultant. These included modeling; prompting; establishing, altering or renaming a discriminative stimulus; physical-response guidance; noncontingent application of hypothesized reinforcers; task alteration; addition of a positive reinforcer, removal of aversive stimulu, extinction; differential reinforcement of other behavior; and counseling. The proportion of time spent on each of the preceding practices was computed. Consultants who used a variety of these procedures in plan implementation received higher scores on flexibility than those who did not. In the multiple regression analysis, flexibility of the consultant contributed significantly to problem identification ($r + .45$).

Four content categories were specified as relevant in problem-identification and problem-analysis interviews: client behavior, problem conditions, recording procedures, and plans for implementation. Content relevance was defined as the amount of information expressed in these categories. In other words, confining the discussion to these four categories was assumed to be an effective consultation practice.

The focus of the interview was measured by an index of content. Consultants who stayed on the same topic for at least five statements were judged more effective than those who switched topics more frequently.

An index of effective verbal processes included verbalizations specifying information, summaries of previously discussed topics, and validations to establish agreement regarding matters of fact. These verbal processes were chosen because in problem solving, for example, the consultant must attain the (1) specification of behavior, (2) the conditions under which the behavior of concern tends to occur, and (3) the specification of a plan to achieve problem solution. In addition, periodic summary statements restating conversation about the problem assists both participants in the consultation process to recall what has been said, while a call for validation assesses the accuracy of the recollections.

Message control referred to the degree to which the consultant guides the problem-solving process. All consultant verbalizations were divided into two categories depending on whether they requested action or information, or provided information. The first, called elicitors, were always interrogative, imperative or exhortive statements. Declarative and exclamatory statements gave information and were called emitters. The overall proportion of consultant elicitors was used as an index of message control. Effective problem solving required the consultant to guide the process by using many elicitors, not emitters. Message control contributed significantly to the multiple correlation.

It was important to note in this study that when a consultant lacked skills or was ineffective, s/he followed some course of action other than problem identification. Problem-solving was then not likely even to be initiated. On the other hand, effective consultants influenced the success of problem-solving by skillfully identifying the problem. When plan implementation occurred, 97% of the cases were satisfactorily resolved.

Teachers with problems were the subjects in a multiple-intervention experimental study. Three elementary teachers demonstrated a reduction in observable negative teacher behavior that Meyers, Freidman and Gaughan (1975) attributed in part to consultation. The techniques used by the counselors included verbal reinforcement, clarification, empathy, direct confrontation, indirect confrontation, probe for feelings, and providing choice. Comparison of total baseline data to total treatment showed Teacher 1 decreased negative behavior from a mean of 5.47% to 2.63%; Teacher 2 decreased from 7.75% to 2.58%; and Teacher 3 decreased from 5.39% to 1.77%. These changes indicate that these techniques may have had an effect.

If we extrapolate from a study by Ivancevich (1982) of the reactions of subordinates to performance appraisal, we can say that feedback plus goal setting are effective components of consultant behavior. The purpose of the study by Ivancevich was to improve the performance appraisal interview skills of team leaders. He trained sixty team leaders, professionals and the employees of a large company. These leaders were divided into four groups of fifteen each and assigned to four different consulting conditions: feedback alone, feedback plus assigned goal setting, goal setting alone, and control. The dependent variables were the reactions of subordinates (consultees) as recorded on a questionnaire, the items of which were classified under the terms equity, accuracy, clarity, comprehensiveness, and motivational impact.

The study indicates that feedback, feedback plus assigned goal setting, and goal setting were better than the control group in the subordinate's perception of equity of the appraisal interview. With respect to accuracy and clarity, feedback and assigned goal setting were superior to the other groups. Motivational impact was not significant in any group. According to this study, if we wish the consultee to feel that the consultation is clear and accurate, we should use feedback and assigned goal setting. However, assigned goal setting intensifies anxiety. The findings of this study are apparently consistent with the work of Kim and Hamner (1976), Erez (1977), Latham, Mitchell, and Dossett (1978), and Nemeroff and Cosentino (1979). It seems clear from this study that consultants are more effective if they are trained to provide feedback and assigned goal setting.

Moreover, a study by Diphoye and de Pontbriand (1981) suggest that employees are likely to accept negative feedback if they are allowed to participate in the feedback session, plans and objectives are discussed, and the factors on which they are evaluated are job related. While negative feelings about negative feedback may not be eliminated entirely by these components of the consultive process, they can be ameliorated by skillful handling. This study involved, as subjects, 971 employees with one and one half or more years of service -- scientists, engineers, or technicians engaged in some aspect of research and development. In another study, Kerr (1976) investigated the use of feedback data, within a clinical supervisory model, to facilitate teachers' selection, implementation, and evaluation of individualized instructional strategies. The study was conducted during a 15 week period in a rural, New England school system. The participants were four self-contained elementary classroom teachers, including a teaching principal. The researcher served as the supervisor and used published instruments to provide feedback to the teachers. The analysis of the data showed that the use of feedback data together with teacher and student completed instruments helped the teacher to select elements of instruction by becoming aware of their instructional processes. Feedback data, critical of the instructional process, when accompanied with suggestions about the individual's instruction, helped the teacher to refine and use individualized teaching strategies.

In an experimental study by Cossairt, Hall & Hopkins (1973) it was hoped to find specific ways principals and consultants could effect positive changes in teacher's behavior through giving instructions, feedback, and praise to the teacher. Effectiveness was measured by changes in incidents of teacher praise in the classroom of the students attending behavior. The experimenters had post-session conferences throughout the experiment and gave positive and contingent statements to the teachers about their use of praise. The results showed that giving teachers instructions about praise, or instructions plus feedback, produced inconclusive data, but by providing contingent praise to the teacher the rates of all three elementary teacher significantly changed. Particularly important to busy principals, the increase were maintained and even ascended when the teachers were placed on an intermittent schedule of praise by the consultants.

Several studies have demonstrated that teacher attention can be used to modify a variety of student behaviors (Cossairt, Hall & Hopkins, 1973; Hall, Lund & Jackson, 1968; Thomas, Becker & Armstrong, 1968), however, the techniques which the consultant must use to modify the teacher's behavior are not as well documented. Meyers, (1975) not only studied the changes in the student behavior, but he attributed behavior changes in highly disruptive third grade black urban children to changes in attitude and behavior of their first year teacher who sought consultation on classroom management. The consultant focussed on changing the attitude of the teacher about being an authority figure. He observed her classroom performance and noted that she began to gain more control of herself, yelled less and found that a word was enough to inhibit disruptive behavior. She expressed more positive feelings about taking control of the classroom, noting that she could still find appropriate times to praise the children and show that she cared about them. The class responded with less inappropriate talking, more in-seat behavior, fewer fights, and they listened to the teacher. Although his study originally dealt with helping a teacher change student behavior through behavior modification techniques, he learned to be a more effective consultant by focussing on the teacher.

Just as we found specific, rather than general praise, to be an important behavioral indicator in effecting study behavior in the classroom and teacher behavior following consultation, the study by Blumberg and Cusick (1970) supports the need to provide specific, rather than general clarifying responses to promote a real sense of understanding of the teacher. The short responses taught to counselors such as "uh huh" or "I see" should be extended to clarify or demonstrate understanding. "What you're saying is that the behavior you see listed on the instrument seems to conflict with your ideas on a child-centered environment. Is that right?" This way of clarifying indicates to the teacher that the consultant is paying attention and understands the problem. Another type of clarifying statement is the reflective re-statement of what the teacher says such as "You don't really feel that these data reflect your true ability."

In the Blumberg and Cusick study only 10% of the time is spent reflecting or clarifying ideas. The teachers viewed the conferences as a waste of time because the consultants were not a source of help. The authors concluded that consultant skills were inadequate, perhaps from lack of training. The teacher must try out consultative plans in the classroom. Another notable variable from this study was the consultant's directive behavior which accounted for .65. Approximately 7 times as much time was given to telling the teacher what to do while less than 1% of the consultant's time was used to ask teachers for suggestions; overall a ratio of .28 for telling to asking. Much time was spent inducing a positive social emotional climate (50%) which apparently was ineffective since teachers rejected the consultation as noncollaborative. Two other findings were (1) consultants received the most negative criticism from asking for information; teachers reacted defensively, and (2) consultants did not handle teacher negativism and tension (-3%). For example, legitimate teacher complaints were handled by agreeing with his/her feelings not by resolving the issues, etc.

This study of taped conferences between teachers and consultants analyzed 15 categories of problem-solving interactional behavior. At issue was helping teachers improve their teaching behavior, but we need to know more about why this was done, i.e., voluntarily, for merit pay, for evaluation or certification.

Although the measure of consultation effectiveness is the subjective judgment of the client, the results of a study by Robbins, Spencer and Frank (1970) are congruent with other studies in that consultant praise ($r = .41$) as supportive behavior was found to be effective. Other variables related positively to client satisfaction were consultant preparation, consultant interest and client's feelings of familiarity toward consultant. The data did deal with program planning and should be viewed cautiously until verified for application to personal consultation.

What the consultant says strongly influences the teacher's response. Tombari and Bergan (1978) studied consultant cues to see if different conceptions of child behavior were activated by the specific diagnostic or behavioral questions asked by consultants about a child's behavior. Teachers given behavioral cues expressed greater optimism about being able to help the child in the classroom than teachers given diagnostic-label cues. This expectancy variable strongly affected problem definition which is the key factor in consultation effectiveness. Thus, this study supports behavioral description of problem behavior and the consultant's cues are instrumental in focussing the discussion on behavior.

8.3 Concept: Problem Formulation

Definition: Problem formulation = consultant behavior that advises or involves the teacher in identifying changes in performance which would eliminate differences between the observed and desired performance.

Indicators

8.3.1 Examines Observation Data

Definition: Examines observation data = consultant behavior that involves the teacher in a cooperative examination of the data coded on the observation instrument in order to identify effective and ineffective performances or omissions of performance.

Example (Pos): Peer Teacher: Why don't you look over the observation data and, for starters, pick out what seems to be the three strongest areas and the three weakest areas. We'll start by reviewing those areas and then proceed to the others, and ultimately to formulating a plan for improvement.

Example (Neg): Supervisor: We don't need any observation data to know the areas where you are ineffective. You are sending too many kids to the office. Why?

8.3.2 Elicits Teacher's Concerns about Data

Definition: Elicits teacher's concerns about data = consultant questions or statements that induce the teacher to talk about the strengths and weaknesses of his or her performance (based on observation data).

Example (Pos): After the teacher and consultant list strength and weaknesses the consultant says: This is a good starting point. Let's try to analyze what it is exactly that makes these strong (or weak) areas for you. For example, you have management of student conduct listed first. What information on the observation sheet indicates to you that this is a weak area?

8.3.3 Seeks Clarification of Concerns

Definition: Seeks clarification of concerns = consultant tests his or her understanding of concerns by repeating teacher's statement(s) or by giving interpretations of what teacher said.

Example (Pos): Consultant: So one of the things you're concerned about is how to stop misbehavior without interrupting the lesson. Is that correct?

8.3.4 Attends to Unnoticed Data

Definition: Attends to unnoticed data = consultant questions or statements that call attention to performance, especially ineffective performance and/or indicators for which there was no performance.

Example (Pos): Consultant: We've discussed most of the indicators on the observation list. However, there are a few we haven't touched on. The observation sheet does not show any tallies for this indicator, number 10, circulates and assists students. Why do you think that is? (Pause.)

8.3.5 Elicits Summary of Data Analysis

Definition: Elicits summary of data analysis = consultant questions or statements that induce the teacher to summarize the results of the data analysis.

Example (Pos): Consultant: Now that we've reviewed the data, how would you summarize your performance? What are your strongest areas and which are the areas we need to look at in depth?

8.3.6 Restates Summary

Definition: Restates summary = consultant states summary of analysis or asks questions to establish agreement regarding the performance to be targeted for change or developed outright.

Example (Pos): Consultant: Let me make sure I have this down. We're going to target question-asking and connected discourse for improvement first. Then we'll work on the use of specific praise, right? Okay, next we will need to develop plans for working on these indicators.

Principle

If the consultant elicits an analysis of the data and retains a problem focus, then the teacher and consultant are more likely to identify the problem.

Supportive Evidence

In a study by Bergan and Tombari (1976) involving eleven psychologists, trained in the application of psychological principles and interviewing skills, with the teachers of 806 children, the most impressive finding was that the skills and knowledge of the consultants had the greatest impact on the problem formulation phase of the problem solving process. If the consultants lack adequate skill and/or ability to apply knowledge, there was a strong likelihood that the problem solving would not occur.

By multiple-regression analysis they found that "indices of interviewing effectiveness, along with measures of consultation skill in applying psychological principles and efficiency in rendering services, accounted for nearly forty-one percent of the variation in whether problem identification occurred." They then go on to say that "the biggest predictor of planned implementation was problem identification. Problem identification accounts for almost fifty-nine percent of the variation in the occurence of planned implementation. Planned implementation in turn accounted for a whopping ninety-five percent of the variation in problem solving." It seems reasonable to extrapolate from the findings of this study that principals, supervisors, or peer teachers who lack interpersonal skills or knowledge of pedagogical research will follow some inefficient course of action in which the teachers will receive little, if any, genuine help.

The training program in behavioral consultation for prospective school psychologists at the University of Arizona included four stages of consultation: problem identification, problem analysis, plan implementation and problem evaluation. In addition to acquiring the knowledge base for analytic problem solving, skills in emitting appropriate verbalizations were demonstrated by trainees through role playing and practicum.

The majority of verbalizations from trained behavioral consultants were classified in four categories: (1) behavior, (2) behavior setting involving antecedent and consequent conditions associated with behavior, (3) observation and recording of behavior, and (4) the establishment of intervention plans and three process categories: (a) elicit specification of information from the consultees, (b) summarize, and (c) validate what has been said.

Evaluation was assessed by the frequency count of appropriate verbalizations by the trainee during each of the four stages, e.g., during problem identification the consultant's statements or questions should cause the consultee to specify behavior of concern to him/her. This coding system provides a way to determine if appropriate verbalizations occur and if requested verbalizations occur during consultation. According to Bergan, Krathochwill, and Luiten (1980) this consultation analysis has been used in numerous studies which produced evidence of its reliability.

In an attempt to empirically test whether consultation is an effective technique for assisting teachers in solving problems within the classroom, Curtis and Watson (1980) found significantly greater improvement in problem-solving skills of teachers who were assigned to highly-skilled consultants than those assigned to low-skilled consultants or those receiving no consultative services. The problem-identification process identified by Bergan and Tombari (1975) as a critical phase in problem-solving were the skills that the authors measured to determine changes in the behavior of 24 elementary and middle school teachers for consultation.

Not only did the quantity of information gather in the problem-identification/clarification interview improve after consultation with high-skilled counselors but the quality also improved because the teachers spent more time describing the problem on the basis of factual data rather than on speculations.

Another important finding in this study indicated that low-skill consultants may have moved teachers too rapidly into planning strategies for problem resolution which is of particular concern in view of the Bergan and Tombari (1976) study in which the key determinant of the effective implementation of plans is successful problem identification.

In a controlled study of the problem identification state of consultation, Medway and Forman (1980) found that teachers preferred direct questioning and responding, specificity, and the use of behavioral principles by consultants.

8.4 Concept: Development of Plans

Definition: Development of plans = consultant behavior that engages the teacher in the formulation of plans to improve performance of the targeted indicators.

Indicators

8.4.1 Elicits Analysis of Targeted Indicators

Definition: Elicits analysis of targeted indicators = consultant questions or statements that invite the teacher to express his or her understanding of the indicators.

Example (Pos): Consultant: You said that you continually have difficulty getting the class to finish the classwork you assign. What are the ways we could check first to determine if the problem lies in the way you handle the class or the clarity of the explanation or the amount of work you expect them to do?

8.4.2 Checks for Agreement

Definition: Checks for agreement = consultant states his or her interpretation of the indicators or restates what the teacher said in order to establish agreement between consultant and teacher.

Example (Pos): Consultant: According to what you said, an effective desist should focus on the correct person, stop the misbehavior from spreading, not interrupt the lesson flow, and not be harsh in nature. Did I leave anything out?

8.4.3 Initiates Planning

Definition: Indicates planning = consultant makes statements or asks questions that stimulate the teacher to join in suggesting ways to improve performance of indicators.

Example (Pos): Consultant: Now that we're both clear on what indicators we want to improve and the behaviors that make up those indicators, the next step is to plan a program for improvement. I have shown you a number of modules and films for self-study that are available in our resource library. We have talked about ways you can record your behavior day by day. You asked about observing me and other teachers. What other ways would you use to learn about these behaviors?

8.4.4 Summarizes Plans

Definition: Summarizes plans = consultant asks questions or makes statements to induce teacher to summarize plans.

Example (Pos): Consultant: I think we've come up with some excellent ideas. To make sure we both understand what the improvement plan will be, let's briefly review the plan and write it down, indicating what each of our responsibilities is. So, for tomorrow what's the first step?

8.4.5 Confirms Plans

Definition: Confirms plans = consultant makes statements or asks questions to test agreement on plans, including materials and activities, time for next observation and interim conference.

Example (Pos): Consultant: Are we agreed that we plan to concentrate on getting control of the classroom again even if some measures seem in conflict with your belief in student-centered operation? You said you would try to explain the need to change classroom procedures that are causing disruptions and elicit the student's help in developing a few rules that all of you could accept. You plan to closely monitor the rules and provide correction for any infraction. Is that correct? (Pause.) We plan to meet briefly at 3:15 p.m. to discuss any difficulties.

Principle

If the consultant engages the teacher in the formulation of a plan for improvement of teacher performance, does not move into planning before the problem is clear, and plans deliberately, not hastily, then changes in teacher performance are likely to occur.

Note: The supportive evidence bearing on planning is found in the preceding section, 8.3.

APPENDIX A

Is Consultation Effective?

The research on consultation is in its infancy. Its findings are scarce and fragile and the dependent variables are more often concerned with attitudes of the consultee toward the consulting situation than with changes in the performance of the consultee or changes in his or her students. Furthermore, much of the research has been done in industry, mental health, and with non-pedagogical personnel such as nurses and police. A considerable amount of research has been done on the school psychologists in the consulting role and some studies have been made of the effects of clinical supervison. These studies may be summarized as follows:

1. To find out which form of consultation--expert or directive consultation, collaborative or participative consultation, or some combination of the two--is more effective.

2. To study the reaction of consultees to the various consulting approaches.

3. To ascertain the effective consulting knowledge and skills.

We shall consider first the effect of consultation on teacher behavior. Does consultation really help the teacher, particularly the teacher who has problems? Mannino and Shore (1975) reviewed 35 empirical studies of the effects of consultation. In 14 of these studies the consultee was the focus of change and 9 focussed on the consultee and either the client or the client and the system, making a total 23 studies dealing with the consultee either singly or in combination. The subjects of these studies were school personnel, police, clergy and nurses. In 17 of the 23 (approximately 74 percent) the change was positive and statistically significant.

Furthermore, in a meta-analysis study, the results of nearly 400 controlled evaluations of therapy demonstrate the benefits of counseling and psychotherapy but few differences in effectiveness were noted among ten different types of treatment. On the average, the typical client is better off than 75% of untreated individuals (Smith and Glass, 1977).

Medway (1979) reviewed 29 studies on the effectiveness of school consultation published between 1972 and 1977, using the following definition as a basis: school consultation is a collaborative problem-solving between a psychologist (the consultant) and a teacher (the consultee), who is responsible for providing some form of assistance to the student (the client). These studies focussed on indirect services and improving the skills of the teacher. In the studies reviewed, consultation was at least partially successful 76% of the time. Of those, 8 studies (28% of total) reported positive results on all dependent measures. Six of these eight studies involved behavioral consultation.

The author states these results should be viewed cautiously because of publisher bias toward studies with favorable results, and because of design flaws common to several studies. Some of the more serious flaws mentioned are lack of a control group, use of an inappropriate control group, not controlling for the number and characteristics of consultation participants, poor evaluation data collection methods, and the focus on attributes of the participants rather than the interaction between consultant and consultee.

Gutkin, Singer and Brown (1980) found teachers had a significant preference for school consultation services over referral services for student problems in three categories of behavior: acting out, withdrawal and academic. Consultants worked with the teachers in a collaborative problem solving process. In an investigation of whether the provision of consultation services affected teacher's perceptions of the severity of student problems, the teachers who worked with consultants viewed the problems listed in Acting Out and Academic subscales as significantly less severe than did the teachers in the control group. The authors note that their results may indicate that "consultation serves to increase the range of problems with which consultees feel they can cope."

APPENDIX B

Forms of Consultation: Direct and Indirect

According to counseling literature, consultation is problem-solving by someone functioning in a guiding role as the consultant. The consultant provides a service, but he/she cannot be totally responsible for the whole process according to those who support the collaborative approach. To reduce the consultee's resistance to a suggested change in his/her behavior the consultant uses counseling techniques such as joint decision making and being nondirective. Because the consultant uses verbal persuasion to induce a behavioral or attitudinal change in a consultee there may be a reaction against the consultant's suggestion if the consultee perceives a threat to his/her freedom. In studies of college subjects Wicklund (1974) compared reactance by subjects who stated a position and received strong support by the consultant, with little reactance by subjects who received a contrary suggestion from the consultant. In the latter case the subjects had explored the pros of the position with which they disagreed so they did not change toward that position; they had demonstrated their freedom to take a position and were not threatened. If the consultant strongly advocated a specific intervention without this collaboration it was likely the consultee would disobey the instructions. Furthermore, collaborative problem-solving may enhance motivation for behavior change. Two categories of democratic, growth-promoting behavior, supervisor accepts or used teacher's ideas and supervisor asks for teacher's opinions, emerged as significant verbal behavior in Reavis' study of clinical supervision (1978). He proposed that these behaviors may be highly significant in promoting teacher motivation for classroom behavior change.

A study of experienced workers by French, Kay and Meyer (1966) indicates that the collaborative mode is superior. A study by Wenger (1979) lends support to the findings of French, Kay, and Meyer. He experimented with two modes of consultation: a) expert approach in which the consultant developed the recommendations for the teacher to follow, and b) collaborative approach in which the consultant involved the teacher in the task of diagnosing the child's educational needs and procedures and techniques for helping the child. Four teachers were the subjects in the collaboative mode and three in the expert mode . The experimenter was the consultant in both cases. The collaborative approach was superior to the expert approach in a number of ways. The teachers were more satisfied; the consultant was perceived as more helpful; the consultant was rated higher on attentiveness to the teacher and on the development of recommendations more appropriate to the teacher's situation.

However, the two appproaches did not differ significantly on the extent to which the recommendations were implemented. In a subsample of teachers of four years experience or less, who received collaborative consultation, it was found that there was a significantly greater follow-through on recommendations than with the other teachers. What this means is not entirely clear, for the degree of influence does not appear comparable to the experiences of beginning teachers.

The data from Blumberg's (1968) study of 210 inservice teachers supported earlier stuies of the perceived quality of interpersonal relationships (positive vs. negative) as related to direct or indirect styles of supervisory behavior (Blumberg and Amidon, 1965; Likert, 1961).

Teachers favored supervisors who emphasized asking for information or opinions and reflecting ideas. Supervisors who were passive or who emphasized the telling dimension but did not reflect etc.. were perceived as less positive or even negative in interpersonal relationships which included empathy. congruence. regard and unconditionality of regard.

Shinn (1976) also reported preferences for non-directive supervisor behavior. In contrast. studies such as Young and Heichberger (1975) reported results of a survey of teachers which indicated a preference for helper behavior defined as essentially directive in nature.

Some studies indicate that collaborative or participative consultation is not likely to be effective with beginners. A study by Hillery and Wexley (1974). involving sixteen undergraduate education majors who taught in a microteaching situation. indicates that the nonparticipative consulting situation is more effective than collaborative situations in producing teaching improvements and in teacher satisfaction with the consultant.

Copeland and Atkinson (1978) explored student teacher preference for supervisory approaches during a year long elementary teacher credential program. Each subject rated supervisors according to eight different concepts; expertness. comprehension of the problem. confidentiality. trustworthiness. ability to help. sincerity. knowledge of training. and utility.

In the audio-tape depicting directive supervision. the supervisor used declarative statements to offer new information asked few questions. and did not consistently reflect or utilize the supervisee's skills. By contrast. the non-directive supervisor made few declarative statements. drew information into the conversation by asking questions and reflected and used the supervisee's ideas.

The differences between the subjects' ratings for the two approaches were statistically significant except for confidentiality. The group mean of the student teachers favored the directive over the non-directive supervisory approach.

Copeland (1980) conducted a similar study of secondary education student teachers. As in the first study. statistical analysis revealed preference for the directive over the non-directive supervisory approach.

A third study was conducted by Copeland (1982) to answer the question of whether student teacher preference would change over time. Preference of supervisory approaches were measured in October and May. The findings of this study led Copeland to suggest that 1) individual preferences differ. and 2) a progression for a mean preference for the directive approach in the Fall moved to a more non-directive approach in the Spring.

Garman (1973) conducted a study in which five college English teaching assistants were given a 12 week teaching seminar in conjunction with clinical supervision. Five other teaching assistants were exposed to the seminar. but

not clinical supervision. The findings indicated that four of the five teaching assistants who received clinical supervision were able to implement the teaching behaviors discussed in the seminar. Only one of the five who did not receive clinical supervision was able to implement the behaviors.

In other studies, more specific elements of clinical supervision were analyzed. Zonca (1972) explored the effects of openness in clinical supervision relationship. Openness was defined as a condition which includes the traits of disclosure, directness, and honesty. These terms were defined as:

disclosure—sharing, revealing, divulging one's thoughts, procedures, and feelings.

directness—straightforward, not vaguely, frankly expressing one's thoughts, procedures, and feelings.

honesty— sincerely, fairly, truthfully expressing thoughts, procedures, and feelings.

The subjects of the study were one intern and one master teacher. The daily logs as well as audio-taped supervisory conferences of the intern and the master teachers were analyzed. From the findings Zonca concluded that the condition of openness contributed in part to the intern's overall development. It also had a positive effect on the ability of the intern to: 1) analyze classroom teaching behaviors, 2) conduct self-evaluations, and 3) become more open with the master teacher.

This exploratory review of the literature offers some support to direct consultation for beginning teachers. However, the literature is limited and the findings inconclusive, thus preventing any conclusions to be drawn about models of effective conferences with beginners.

REFERENCES

Bergan, J. R. Behavioral Consultation. Columbus, Ohio: Charles E. Merrill (1977).

Bergan, J. R.; Kratochwill, T. R.; and Luiten, J. Competency-Based Training in Behavioral Consultation. Journal of School Psychology (1980) 18:91-97.

Bergan, J. R. and Tombari, M. L. Consultant Skill and Efficiency and the Implementation and Outcomes of Consultation. Journal of School Psychology (1976) 14:3-14.

Bergan, J. R. and Tombari, M. L. The Analysis of Verbal Interactions Occurring during Consultation. Journal of School Psychology (1975) 13:209-226.

Blumberg, A. Supervisory Behavior and Interpersonal Relations. Education Administration Quarterly (Spring 1968) 7:34-45.

Blumberg, A. and Amidon, E. Teacher Perceptions of Supervisor-Teacher Interaction. Administrator's Notebook (1965) 14:1-8.

Blumberg, A. and Cusick, P. Supervisor-Teacher Interaction: An Analysis of Verbal Behavior. Education (1970) 91:126-134.

Bolster, B. I. and Springbett, B. M. The Reaction of Interviewers to Favorable and Unfavorable Information. Journal of Applied Psychology (1961) 45:97-103.

Broekmann, N. C. and Moller, A. T. Preferred Seating Position and Distance in Various Situations. Journal of Counseling Psychology (1973) 20:504-508.

Caplan, G. The Theory and Practice of Mental Health Consultation. New York: Basic Books, 1970, 308-327.

Copeland, W. D. Affective Dispositions of Teachers in Training toward Examples of Supervisory Behavior. Journal of Educational Research (1980) 74:37-42.

Copeland, W. D. Student Teachers' Preference for Supervisory Approach. Journal of Teacher Education (1982) 33:32-36.

Copeland, W. D. and Atkinson, D. R. Student Teachers' Perceptions of Directive and Non-Directive Supervisory Behavior. Journal of Educational Research (1978) 71:123-127.

Cossairt, A.; Hall, R. V.; and Hopkins, B. L. The Effects of Experimenter's Instructions, Feedback, and Praise on Teacher Praise and Student Attending Behavior. Journal of Applied Behavior Analysis (1973) 6:69-100.

Curtis, M. J. and Watson, K. L. Changes in Consultee Problem Clarification Skills Following Consultation. Journal of School Psychology (1980) 18:3, 210-221.

Diphoye, R. L. and dePontbriand, R. Correlates of Employee Reactions to Performance Appraisals and Appraisal Systems. Journal of Applied Psychology (1981) 66:248-251.

Erez, M. Feedback: A Necessary Condition of the Goal Setting-Performance Relationship. Journal of Applied Psychology (1977) 62:624-627.

French, J. R. P.; Kay, E.; and Meyer, H. H. Participation and the Appraisal System. Human Relations (1966) 19:3-20.

Garman, N. B. A Study of Clinical Supervision as a Resource of College Teachers of English. Dissertation Abstracts International (1971) 32:6835A.

Gottschalk, L. A. and Gleser, G. C. The Measurement of Psychological States Through the Content Analysis of Verbal Behavior, Berkeley, CA.: University of California Press, 1969.

Gottschalk, L. A.; Winget, C. N; and Gleser, G. C. Manual of Instructions for Using the Gottschalk-Gleser Content Analysis Scales. Berkeley, CA.: University of California Press, 1969.

Greene, L. R. Effects of Verbal Evaluative Feedback and Interpersonal Distance on Behavioral Compliance. Journal of Counseling Psychology (1977) 25:10-14.

Gutkin, T. B.; Singer, J. H.; and Brown, R. Teacher Reaction to School-Based Consultation Services: A Multivariate to Analysis. Journal of School Psychology (1980) 18:126-134.

Haase, R. F. and DiMattia, D. J. Proxemic Behavior: Counselor, Administrator, and Client Preference for Seating Arrangement in Dyadic Interactia. Journal of Counseling Psychology (1976) 17:319-325.

Hall, R. V.; Lund, D.; and Jackson, D. Effects of Teacher Attention on Study Behavior. Journal of Applied Behavior Analysis (1968) 1:1-22.

Hillery, J. M. and Wexley, K. N. Participation Effects in Appraisal Interviews Conducted in a Training Situation. Journal of Applied Psychology (1974) 59:168-171.

Holloway, E. L. Interactional Structure of the Supervision Interview. Journal of Counseling Psychology (1982) 29:309-317.

Hughes, J. N. and Falk, R. S. Resistance, Reactance, and Consultation. Journal of School Psychology (1981) 19:134-142.

Ivancevich, J. M. Subordinates' Reactions to Performance Appraisal Interviews: A Test of Feedback and Goal Setting Techniques. Journal of Applied Psychology (1980) 67:581-587.

Jason, L. A. and Ferone, L. Behavioral Versus Process Consultation Interventions in School Settings. American Journal of Community Psychology (1978) 6:531-543.

Kerr, R. G. The Relationship among Attitude Scores, Dogmatism Scores, and Change in a Classroom Teaching Pattern of Teachers Who have Experienced Clinical Supervision. Dissertation Abstracts International (1976) 37:2132A.

Keutzer, C. S.; Fosmire, F. R.; Diller, R.; and Smith, M. D. Laboratory Training in a New Social System: Evaluation of a Consulting Relationship with a High School Faculty. Journal of Applied Behavioral Science (1971) 7:493-501.

Kim, J. S. and Hamner, W. C. Effect of Performance Feedback and Goal Setting on Productivity and Satisfaction in an Organizational Setting. Journal of Applied Psychology (1976) 61:48-57.

Latham, G. P.; Mitchell, T. R.; and Dossett, D. L. Importance of Participative Goal Setting and Anticipated Rewards on Goal Difficulty and Job Performance. Journal of Applied Psychology (April 1978) 63:163-171.

Likert, R. New Patterns of Management. New York: McGraw Hill, 1961. p. 103.

Likert, R. Motivational Approach to Management Development. Harvard Business Review (1959) 37:75-82.

Mannino, F. V. Task Accomplishment and Consultation Outcome. Community Mental Health Journal (1972) 8:102-108.

Mannino, F. V. and Shore, M. F. The Effects of Consultation: A Review of Empirical Studies. American Journal of Community Psychology (March 1975) 3:1-21.

Medway, F. J. How Effective is School Consultation? A Review of Recent Research. Journal of School Psychology (1979) 17:275-282.

Medway, F. J. and Forman, S. G. Psychologists' and Teachers' Reactions to Mental Health and Behavioral School Consultation. Journal of School Psychology (1980) 18:338-348.

Meyers, J. A Consultation Model for School Psychological Services. Journal of School Psychology (1973) 11:5-15.

Meyers, J. Consultee-Centered Consultation with a Teacher as a Technique in Behavior Management. American Journal of Community Psychology (June 1975) 3:111-121.

Meyers, J.; Freidman, M. P.; and Gaugham, E. J. The Effects of Consultee-Centered Consultation on Teacher Behavior. Psychology in the Schools (1975) 12:288-295.

Meyers, J.; Freidman, M. P.; Gaughan, E. J.; and Pitt, N. An Approach to Investigate Anxiety and Hostility in Consultee-Centered Consultation. Psychology in the Schools (April 1978) 15:292-296.

Meyers, J.; Pitt, N. W.; Gaughan, E. J.; and Freidman, M. P. A Research Model for Consultation with Teachers. Journal of School Psychology (1978) 16:137-145.

Nemeroff, W. F. and Cosentino, J. Utilizing Feedback and Goal Setting to Increase Performance Appraisal Interview Skills of Managers. Academy of Management Journal (1979) 22:566-576.

Reavis, C. A Test of the Clinical Supervision Model. Journal of Educational Research (July-August, 1977) 70:311-315.

Reavis, Charles A. Research in Review/Clinical Supervision: A Review of the Research. Educational Leadership (April 1978) 580-584.

Robbins, P. R. and Spencer, E. C. A Study of the Consultation Process. Psychiatry (1968) 31:362-368.

Robbins, P. R.; Spencer, E. C.; and Frank, D. A. Some Factors Influencing the Outcome of Consultation. American Journal of Public Health (March 1970) 60:524-534.

Schmuck, R. A. Helping Teachers Improve Classroom Group Processes. Journal of Applied Behavioral Science. (1968) 4:401-435.

Shinn, J. L. Teacher Perceptions of Ideal and Actual Supervisory Procedures Used by California Elementary Principals: The Effects of Supervisory Training Programs Sponsored by the Association of California School Administrators. Dissertation Abstracts International (1976).

Shuma, K. Y. Changes Effectuated by a Clinical Supervisory Relationship Which Emphasize a Helping Relationship and a Conference Format Made Congruent with the Establishment and Maintenance of This Helping Relationship. Dissertation Abstracts International (1973) 35:729A.

Smith, M. L. and Glass, G. V. Meta-Analysis of Psychotherapy Outcome Studies. American Psychologist (September 1977) 752-760.

Thomas, D. R.; Becker, W. C.; and Armstrong, M. Production and Elimination of Disruptive Classroom Behavior By Systematically Varying Teacher's Behavior. Journal of Applied Behavior Analysis (1968) 1:35-45.

Tombari, M. L. and Bergan, J. R. Consultant Cues and Teacher Verbalizations, Judgments, and Expectancies Concerning Children's Adjustment Problems. Journal of School Psychology (1978) 16: 212-219.

Wenger, R. D. Teacher Response to Collaborative Consultation. Psychology in the Schools (January 1979) 16:127-131.

Wicklund, R. A. Freedom and Reactance. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1974.

Young, J. M. and Heichberger, R. L. Teachers' Perceptions of an Effective School Supervision and Evaluation Program. Education (1975) 96:10-19.

Zonca, P. H. A Case Study Exploring the Effects on an Intern Teacher of the Condition of Openness in a Clinical Supervisory Relationship. Dissertation Abstracts International (1972) 33:658A.

APPENDICES SECTION III, A-B

Proposed Feedback System

Section III, Appendix A

OPSCAN FORM

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

Institution _____

Date _____

PRINT LAST NAME FIRST

SPECIAL CODING SECTION

A B C D E F G H I J

IDENTIFICATION NO.

237

122

238

Section III, Appendix B

Beginning Teacher Questionnaire

At which university did you do most of your preparation for your teaching certificate? Find the 4-digit number for the university in the Coding for Florida Schools (see Attachment 1 in this Appendix). Enter that number in columns A, B, C, and D of the Special Coding Section of your answer sheet by darkening the appropriate circles.

In what areas are you certified? Find the 3-digit code for each area in which you are certified in the Master Coding List (see Attachment 2 in this Appendix). Enter the code(s) in the Identification No. Section of your answer sheet. List codes in order of the amount of training you received (greatest to least). Enter no more than three codes. Do not space between codes.

For questions 1-9, darken the appropriate circle in the Answer Section (1-60) of your answer sheet. For example, if you are certified in grades K-6, darken circle #1 under answer #1.

1. Grade levels in which you are certified to teach:

   1. K-6, 2. 7-9, 3. 7-12, 4. 9-12, 5. K-12.

2. Number of years teaching experience:

   1. Less than one year
   2. One to five years
   3. Five or more years

3. Highest degree held:

   1. Non-degree (vocational)
   2. B.A., B.S.
   3. M.A., M.S., E.Ed.
   4. Advanced Graduate

4. Number of semester hours in Education you have taken.

   1. 0, 2. 1-18, 3. 19-30, 4. 31-40, 5. 41 plus.

5. In the subject you are teaching, how many semester hours have you taken outside the College of Education?

   1. 0, 2. 1-18, 3. 19-30, 4. 31-40, 5. 41 plus.

6. Are you teaching courses outside your area of certification? (Either grade level or subject area.)

   1. Yes 2. No

7. If you are an elementary or middle school teacher and answered yes to #6, how much of your teaching time each day is spent outside your area of certification?

   1. 0, 2. 1-15%, 3. 16-31%, 4. 32-50%, 5. 51% plus.

8. If you are a junior high or high school teacher and answered yes to #6, how much of your teaching time each day is spent outside your area of certification?

   1. one period 3. five periods 5. five periods
   2. two periods 4. four periods

9. Were you (will you be) observed teaching outside your area of certification?

   1. Yes 2. No

Questions 10-19 deal with effectiveness of your university training. Rate the preparation you received from your university or college in each area listed. Circle the most appropriate answer, then darken the corresponding circle on your answer sheet.

	Not Taught	Not At All Effective	Somewhat Effective	Very Effective
<u>Preparation for Teaching</u>:				
10. Planning of Instruction	0	1	2	3
11. Control of Student Conduct (discipline)	0	1	2	3

	Not Taught	Not At All Effective	Somewhat Effective	Very Effective
12. Instructional Organization and Development	0	1	2	3
13. Presentation of Subject Matter	0	1	2	3
14. Communication: Verbal and Nonverbal	0	1	2	3
15. Teaching Test Taking Skills, Giving Tests and Test Feedback	0	1	2	3
16. Intern Experience	0	1	2	3

Rate the work you took outside the Department or College of Teacher Preparation:

17. Junior College (if applicable)	0	1	2	3
18. Academic Subjects (Freshmen and Sophomore)	0	1	2	3
19. Academic Subjects (Junior and Senior)	0	1	2	3

For questions 21-22, darken the appropriate circle in the answer section of your answer sheet.

20. What is your sex?

    1. Female 2. Male

21. To what ethnic group do you belong?

    1. White--nonHispanic 2. Black-nonHispanic

    3. Hispanic 4. Asian/Pacific islands

    5. American Indian/Alaskan native

It is assumed that the Beginning Teacher Questionnaire will be administered by the principal or a designee. The following instructions are suggested:

Ask each beginning teacher to complete the attached questionnaire and to

1. Enter his/her name (last name first) on the line marked "name" on the answer sheet.
2. Enter the name of the institution from which he or she received the Bachelor's degree on the line marked "institution," if the teacher is from out-of-state.
3. Answer the questions on the Beginning Teacher Questionnaire, using the accompanying system of codes, recording the answers on the answer sheet.
4. Turn in answer sheets and all material to the administrator.

04/09/84
CPT:A.PFS1-5

Attachment 1

CODING FOR COLLEGES AND UNIVERSITIES

5053	Barry College
5061	Bethune-Cookman
5076	Biscayne College
5142	Clearwater Christian
5223	Eckerd College
5182	Edward Waters College
5190	Embry-Riddle Aero University
5235	Flagler College
5215	Florida A&M University
5229	Florida Atlantic
5080	Florida Institute of Technology
5206	Fla. International University
5217	Fla. Memorial College
5218	Fla. Southern College
5219	Florida State University
5331	Jacksonville University
5506	New College of South Fla.
5514	Nova University
5553	Palm Beach Atlantic
5572	Rollins
5638	St. Leo College
5630	Stetson University
9999	Southeastern Bible College
5233	University of Central Florida
5812	University of Florida
5815	University of Miami
5490	University of North Florida
5828	University of South Florida
5819	University of Tampa
5833	University of West Florida
5883	Warner Southern College
9998	Miami Christian College
5000	Non-Florida College or University

ATTACHMENT 2

MASTER CODING LIST

100 Administration (Adult Education)
101 Administration (K-6)
102 Administration (7-12)
103 Administration (K-12)
104 Administration and Supervision (K-6)
105 Administration and Supervision (7-12)
106 Administration and Supervision (K-12)
107 Administration/Guidance (K-6)
108 Administration/Guidance (7-12)
109 Administration/Guidance (K-12)
110 Administration/Supervision/Emotional Distrurbance (K-12)
111 Administration/Supervision/Hearing (K-12)
112 Administration/Supervision/Mental Retardation (K-12)
113 Administration/Supervision/Specific Learning Disabilities (K-12)
114 Art Education (K-12)
115 Bible (7-12)
116 Biology (7-12)
117 Business Education (7-12)
118 Chemistry (7-12)
119 Distributive Education
120 Drama (7-12)
121 Drama/Speech (7-12)
122 Early Childhood Education (N-K)
123 Early Childhood Education/Elementary Education
124 Early Childhood Education/Emotional Disturbance
125 Early Childhood Education/Hearing
126 Early Childhood Education/Mental Retardation
127 Early Childhood Education/Special Learning Diabilities
128 Earth Science (7-12)
129 Economics (7-12)
130 Educational Media Specialist (K-12)
131 Elementary Education (1-6)
132 Elementary Education/Emotional Disturbance
133 Elementary Education/Hearing
134 Elementary Education/Mental Retardation
135 Elementary Education/Specific Learning Disabilities
136 Elementary Education/Varying Exceptionalities
137 Elementary Education/Emotional Distrubance/Specific Learning Disabilities/Mental Retardation
138 Elementary Education/Elementary Mathematics
139 Elementary/Middle School English (1-MS)
140 Elementary Education/Middle School Mathematics (1-MS)
141 Elementary Education/Middle School Science (1-MS)
142 Elementary Education/Middle School Social Studies (1-MS)

143 Elementary Education/Early Ehildhood Education/Middle School English (K-MS)
144 Elementary Education/Early Childhood Education/Middle School Mathematics (K-MS)
145 Elementary Education/Early Childhood Education/Middle School Science (K-MS)
146 Elementary Education/Early Childhood Education/Middle School Social Studies (K-MS)
147 Emotional Disturbance (K-12)
148 Emotional Disturbance/Guidance (K-12)
149 Emotional Disturbance/Physical Education (K-12)
150 Emotional Disturbance/Reading (K-12)
151 Emotional Disturbance/Mental Retardation/Specific Learning Disabilities (K-12)
152 Emotional Disturbance/Specific Learning Disabilities/Varying Exceptionalities (K-12)
153 Emotional Disturbance/Mental Retardation/Specific Learning Disabilities/Varying Exceptionalities (K-12)
154 English (MS)
155 English (7-9)
156 English (7-12)
157 English/Hearing
158 English/Journalism (7-12)
159 English/Speech (7-12)
160 English/Social Studies (MS; 7-9)
161 French (7-12)
162 Geography (7-12)
163 German (7-12)
164 Gifted (K-12)
165 Guidance (K-12)
166 Guidance/Hearing (K-12)
167 Guidance/Mental Retardation (K-12)
168 Guidance/Specific Learning Disabilities (K-12)
169 Health Education (7-12)
170 Health Occupations
171 Health Occupations PBTE
172 Health/Physical Education
173 Hearing Disabilities (K-12)
174 Hearing Disabilities/Mental Retardation (K-12)
175 Hearing Disabilities/Physical Education (K-12)
176 Hearing Disabilities/Reading (K-12)
177 Hearing Disabilities/Social Studies (7-12)
178 History (7-12)
179 Humanities (K-12)
180 Industrial Arts (7-12)
181 Industrial Education
182 Industrial Education/Technical Education
183 Industrial Education/Technical Education PBTE
184 Industrial Education/Technical Education/Health Occupations
185 Italian (7-12)
186 Journalism (7-12)
187 Latin (7-12)
188 Mathematics (1-6)
189 Mathematics (MS)
190 Mathematics (7-9

191 Mathematics (7-12)
192 Mathematics (1-6/MS)
193 Mathematics (7-9/MS)
194 Mathematics/Science (7-9)
195 Mathematics/Science (7-12)
196 Mathematics/Science (MS, 7-9)
197 Mental Retardation (K-12)
198 Mental Retardation/Physical Education (K-12)
199 Mental Retardation/Reading (K-12)
200 Motor Disabilities (K-12)
201 Motor Disabilities/Physical Education (K-12)
202 Music Education (K-12)
203 Music Education (7-12)
204 Music Instrumental (K-12)
205 Music Vocal (K-12)
206 Physical Education (K-12)
207 Physical Education/Specific Learning Disabilities (K-12)
208 Physics (7-12)
209 Political Science (7-12)
210 Portuguese (7-12)
211 Psychology (7-12)
212 Reading (K-12)
213 Reading/Specific Learning Disabilities (K-12)
214 Russian (7-12)
215 Science (1-6)
216 Science (MS)
217 Science (7-9)
218 Science (7-12)
219 Spanich (7-12)
220 Specialist in School Psychology (K-12)
221 Specific Learning Disabilities (K-12)
222 Speech (7-12)
223 Speech Correction (K-12)
224 Social Studies (MS)
225 Social Studies (7-9)
226 Social Studies (7-12)
227 Sociology (7-12)
228 Supervision (K-6)
229 Supervision (7-12)
230 Supervision (K-12)
231 Technical Education
232 Varying Exceptionalities (K-12)
233 Visiting Teacher (K-12)
234 Visual Disabilities (K-12)
235 Vocational Education Director
236 Vocational Home Economics
237 Vocational Office Education
238 Vocational Agriculture
239 English (7-12)/Spanish (7-12)
240 English/Speech/Journalism
241 Elementary Education/Early Childhood - Add-on
242 Mathematics (7-12)/Mathematics (MS)
243 English (7-12)/French (7-12)
244 Spanish (7-12)/French (7-12)
245 Mathematics (7-12) Add-on
246 Social Studies (7-12) Add-on

500	Bilingual Education
501	ESOL

Appendix Section IV, A

Cognitive Examination

APPENDIX A

Test Item Writers

Charles Ahearn

Jean M. Borg

Mary Bullerman

G. Allen Geyer

Gene Olliff

Donald E. Orlosky

Karen Shapiro

B. Othanel Smith

Nancy Thigpen

Melva Underbakke

Nancy Wilson

APPENDIX

Study Guide for the Cognitive Examination

1. Review the Domain documents 2-6.
2. There are approximately ten questions on each domain.
3. The questions are multiple choice and matching.
4. The questions assess knowledge of the research content, not the identification of the researchers. It is important to know the vocabulary of effective and ineffective pedagogical practice.

Here are some sample questions from Domain 1 which are not included on the present test:

1. The teacher says, "Since this skill is on the state assessment test, I want to cover it before the test is given next week." This is an example of
   a. goal justification
   b. content evaluation
   c. diagnosis
   d. activity structuring
2. Analysis of content involves determining whether the content is
   a. fact or opinion
   b. cognitive or affective
   c. concept, skill, or rule
   d. correct or incorrect

Directions for questions 3-6

Match the named concept with the example:

a. Content coverage

b. Utilization of instructional materials

c. Activity structure

d. Goal focusing

e. Diagnosis

3. _____ "Practice sheet, page 55, deals with changing telling sentences to asking sentences by using a question word."

4. _____ "They will read a play, each one reading a separate part. We will read the play silently, paint a mural, discuss the story. I will introduce the vocabulary words on the board."

5. _____ "With my other groups I did the graduated beads with size and it was okay, but this group is so slow. I'm really going to have to do something just large and small to begin with, and then large, medium, and small."

6. _____ "Definition by context will be discussed."

We will use Trans-Optic answer sheets. Please bring a black lead pencil (No. 2 1/2 or softer).

04/02/84
CPT:A.ASG1-3

Appendices Section V, A-B

A Proposed Revision of the Florida Generic Competencies

Appendix Section V, A

The Original Florida Generic Competencies

The Original Florida Generic Competencies and Subskills

1. Demonstrate the ability to orally communicate information on a given topic in a coherent and logical manner.
   a. Utilizes principles of simplicity and clarity in organization of oral presentation.
   b. Uses standard English in oral communication.
   c. Uses vocabulary suitable to the topic and audience.
   d. Speaks with a volume and pace that promote comprehension.
   e. Provides verbal and nonverbal cues to the organizational structure of the oral message.
   f. Provides relevant examples that illustrate oral content.
2. Demonstrate the ability to write in a logical easily understood style with appropriate grammar and sentence structure.
   a. Differentiates between formal and informal written English and demonstrates ability to use both forms.
   b. Uses language at the level appropriate to the topic and reader.
   c. Comprehends and applies basic mechanics of writing: spelling, capitalization, and punctuation.
   d. Comprehends and applies appropriate sentence structure.
   e. Comprehends and applies basic techniques for the organization of written material.
   f. Comprehends and applies standard English usage in written communication.
3. Demonstrate the ability to comprehend and interpret a message after listening.

a. Accurately follows multi-step oral directions.

b. Listens effectively for the organization, main idea, subordinate ideas and details of a message.

c. Listens effectively in order to identify relevant and irrelevant information and propaganda techniques.

d. Listens effectively in order to draw inferences.

e. Summarizes the message after listening.

f. Comprehends both standard and nonstandard English language used by students.

4. Demonstrate the ability to read, comprehend, and interpret, orally and in writing, professional material.

a. Identifies and evaluates relevant professional material.

b. Understands basic statistical terminology (such as mean, median, mode).

c. Demonstrates literal reading skills (such as recognizing main idea, details, sequencing, comparison and contrast).

d. Demonstrates interpretive reading skills (such as predicting outcome, drawing conclusions, making generalizations).

e. Demonstrates critical reading skills (such as recognition of relevant and irrelevant information, propaganda techniques, and fallacies in reasoning).

f. Produces a logical summary interpretation of the results of research in professional material.

5. Demonstrate the ability to comprehend and work with fundamental mathematical concepts.

a. Adds, subtracts, multiplies and divides whole numbers, decimals and

fractions.

b. Demonstrates the meaning and use of fractions and percents.

c. Represents and interprets data using charts, tables, graphs and maps.

d. Solves measurement problems involving length, area, volume, capacity, weight, time and temperature, using U.S. customary and metric units.

e. Applies mathematical skills to solve real world problems.

f. Identifies geometric forms and relationships.

6/19. The ability to comprehend patterns of physical, social and academic development in students, including exceptional students in the regular classroom, and to counsel the same students concerning their needs in these areas.

a. Demonstrates knowledge of basic principles of human growth and development, and awareness of individual differences in students.

b. Obtains knowledge of students through classroom tests, teacher observations, and student records to contribute to understanding of student needs.

c. Explains test data and other classroom evaluations to students in relation to their educational needs.

d. Applies motivational techniques to encourage students to be achievement-oriented and goal-directed.

e. Assists the student in relating achievements and interests to aptitude and ability.

f. Demonstrates knowledge of alternative school and community resources for students who have special needs.

g. Assists students in developing individual learning activities.

h. Develops student awareness of career opportunities using school and community resources.

7. Diagnose the entry level knowledge and/or skills of students for a given set of instructional objectives using diagnostic tests, teacher observation and student records.

   a. Selects the specified knowledge or skill to be diagnosed, and determines the most appropriate method for conducting the diagnosis.

   b. Selects or constructs a test to diagnose student learning needs.

   c. Uses classroom observation techniques to diagnose student learning needs.

   d. Uses information from student records to diagnose student learning needs.

   e. Interprets results obtained from diagnostic tests, teacher observation and information from student records.

8. Identify long-range goals for a given subject area.

   a. Identifies state and district long-range goals.

   b. Formulates subject area goals consistent with state and district goals and student needs.

9. Construct and sequence related short-range objectives for a given subject area.

   a. Identifies knowledge, skills and attitudes to be attained for a subject area.

   b. Constructs or adapts short-range objectives for identified knowledge, skills and attitudes.

c. Sequences short-range objectives consistent with commonly accepted principles of learning.

10. Select, adapt and/or develop instructional materials for a given set of instructional objectives and student learning needs.
    a. Determines desirable characteristics of materials based on objectives and student learning needs.
    b. Locates and evaluates available instructional material.
    c. Selects materials to assist students in mastering an objective.
    d. Demonstrates techniques for modifying materials to assist students in mastering an objective.
    e. Determines materials to be developed based on existing resources and student needs.
    f. Identifies and selects resources needed for materials development.
    g. Designs and constructs materials based on instructional objectives, student needs and available resources.
    h. After use, evaluates the effectiveness of instructional materials in accomplishing objectives, and revises accordingly.
11. Select/develop and sequence related learning activities appropriate for a given set of instructional objectives and student learning needs.
    a. Comprehends basic principles of human growth and development.
    b. Identifies conditions that affect learning.
    c. Identifies alternative activities to achieve an objective.
    d. Selects an appropriate learning activity to achieve an objective.
    e. Combines appropriate learning activities into an instructional sequence.
12. Establish rapport with students in the classroom by using verbal and/or

visual motivational devices.

a. Secures the attention of students through appropriate techniques.

b. Relates instructional objectives and activities to interests, capabilities and experiences of students.

c. Informs students about objectives, subsequent learning tasks and performance expectations.

d. Explains choices and limitations of possible learning activities.

e. Alters instructional strategies during learning activities based on student responses and other factors.

f. Relates students' and teacher's experiences, thoughts and feelings to learning activities.

g. Uses reinforcement techniques to assist in student motivation.

h. Uses media to secure interest and maintain attention.

i. Uses student products and talent to secure interest and maintain attention.

13. Present directions for carrying out an instructional activity.

a. Selects appropriate means for presenting directions.

b. Secures attention of students for the purpose of giving directions.

c. Informs students of objectives, assessments and performance standards.

d. Informs students of the sequence and nature of learning activities to achieve the objectives.

e. Identifies materials for a learning activity and explains their use.

f. Determines if students understand directions.

g. Clarifies directions by responding to student questions.

14. Construct or assemble a classroom test to measure student performance according to criteria based on objectives.
    a. Identifies uses of basic types of classroom tests and assessment techniques.
    b. Identifies appropriate uses of norm-referenced and criterion-referenced testing.
    c. Given an objective, specifies knowledge and skills to be assessed.
    d. Selects appropriate assessment techniques to evaluate mastery of an objective.
    e. Determines limitations, constraints and requirements for administering tests.
    f. Constructs and identifies test items and tasks that evaluate mastery of an objective.
    g. Identifies criteria for standards of performance.
    h. Assembles test components including test items, directions and scoring keys.
    i. Evaluates and/or revises tests on the basis of validity, reliability and student responses.
15. Establish a set of classroom routines and procedures for utilization and care of materials.
    a. Involves students in developing classroom routines and procedures for utilization and care of materials.
    b. Determines the type and amount of materials necessary to complete classroom assignments.
    c. Organizes an effective system for placement and distribution of materials in the classroom.

d. Organizes and arranges a center that will serve as a focus of interest for student learning (such as bulletin board, display table, or exhibit).

e. Identifies physical elements and arrangements in the classroom that directly affect learning.

f. Involves students in developing routines and procedures for physical movement in the classroom.

g. Arranges classroom furniture and equipment to accommodate selected teaching strategies.

h. Identifies approved procedures for movement of students in emergencies that can be anticipated.

16. Formulate a standard for student behavior in the classroom.

a. Identifies approved safety procedures and incorporates them into a standard for student behavior in the classroom.

b. Identifies and incorporates socially accepted norms (such as mutual respect, consideration of others, courtesy) into a standard for student behavior in the classroom.

c. Identifies characteristics of the student population (such as age and maturity) that need to be considered in formulating a standard for student behavior in the classroom.

d. Establishes a realistic standard of behavior that has potential for consistent application.

e. Identifies and incorporates state and local policies into a standard for student behavior in the classroom.

17. Identify causes of classroom misbehavior and employ a technique(s) for correcting it.

a. Identifies factors of the physical environment that affect student behavior.

b. Identifies social and emotional characteristics of the teacher that affect student behavior.

c. Identifies physical, social and emotional characteristics of the student that affect student behavior.

d. Identifies out-of-school factors that affect student behavior.

e. Identifies aspects of instructional procedures and techniques which affect student behavior.

f. Demonstrates effective techniques and strategies for managing student behavior.

g. Uses selected verbal and nonverbal techniques for enforcing and modifying student behavior.

h. Identifies and uses school and community resources for assistance in modifying student behavior.

i. Obtains and utilizes parental assistance for modifying student behavior.

18. Identify and/or develop a system for keeping records of class and individual student progress.

a. Constructs a system for recording individual student knowledge and skills progress in a subject area.

b. Identifies methods for reporting individual student progress in knowledge and skills in a subject area.

c. Identifies methods for recording class progress in knowledge and skills in a subject area.

d. Identifies methods for reporting class progress in knowledge and

skills in a subject area.

e. Demonstrates knowledge of the laws and policies governing the content and use of student records.

19. (See Section #6)

20. Identify and/or demonstrate behaviors which reflect a feeling for the dignity and worth of other people including those from other ethnic, cultural, linguistic and economic groups.

a. Creates a learning environment in which students express themselves openly and honestly.

b. Assists students in understanding that individual differences enable each person to make unique contributions to the group effort.

c. Demonstrates awareness of cultural differences in dress, beliefs and practices.

d. Establishes an environment for positive communication and interaction between students from different social-cultural backgrounds.

21. Demonstrate instructional and social skills which assist students in developing a positive self-concept.

a. Exhibits behavior in the classroom that is empathetic, positive and reinforcing.

b. Assists students in initiating self-directed learning.

c. Assists students in understanding their needs, motives, experiences and individual value and dignity.

d. Selects and uses curriculum materials in accordance with the abilities and mastery levels of individual students.

22. Demonstrate instructional and social skills which assist students in

interacting constructively with their peers.

a. Establishes an environment that permits students to cooperate and share ideas and materials.

b. Assists students in applying constructive criticism in response to each other's work.

c. Establishes a learning environment designed to assist students in exhibiting positive interpersonal traits (such as mutual respect and cooperation).

d. Uses techniques that assist students in examining their values, attitudes and beliefs.

23. Demonstrate teaching skills which assist students in developing their own values, attitudes and beliefs.

a. Assists students in understanding the need to explore alternative solutions to problems.

b. Establishes teaching strategies that allow students to make choices based on clearly defined consequences.

24. Beginning July 1, 1982, the ability to recognize and be aware of the instructional needs of exceptional students.

a. Identifies the characteristics of exceptional students that have implications for modifying the learning environment.

b. Demonstrates awareness and appropriate use of educational programs, support services, personnel and other resources available to meet the needs of exceptional students.

c. Demonstrates the ability to identify and appropriately refer students who may be in need of exceptional student education.

d. Demonstrates awareness of the roles of the parent, teacher, and

other professional personnel as members of the educational team responsible for planning, implementing, and evaluating the exceptional student's program.

e. Demonstrates the ability to recognize and/or use alternative instructional strategies to implement that portion of the exceptional student's program for which the teacher has responsibility.

f. Identifies and/or selects effective techniques and strategies for facilitating integration and social acceptance of exceptional students.

04/02/84
CPT:A.A01-14

Appendix Section V, B

Proposed Revision of the Florida Generic Competencies

Competencies Most Appropriately Evaluated by Cognitive Tests

1. Demonstrate the ability to write in a logical, easily understood style with appropriate grammar and sentence structure.
   a. Differentiates between formal and informal written English and demonstrates ability to use both forms.
   b. Uses language at the level appropriate to the topic and reader.
   c. Comprehends and applies basic mechanics of writing, spelling, capitalization, and punctuation.
   d. Comprehends and applies appropriate sentence structure.
   e. Comprehends and applies basic techniques for the organization of written materials.
   f. Comprehends and applies standard English usage in written communication.
2. Demonstrate the ability to comprehend and interpret a message after listening.
   a. Accurately follows multi-step oral directions.
   b. Listens effectively for the organization, main ideas, subordinate ideas and details of a message.
   c. Listens effectively in order to identify relevant and irrelevant information and propaganda techniques.
   d. Listens effectively in order to draw inferences.
   e. Summarizes the message after listening.
   f. Comprehends both standard and nonstandard English language used by students.
3. Demonstrate the ability to read, comprehend, and interpret, orally and

in writing, professional material.

a. Identifies and evaluates relevant professional material.

b. Understands basic statistical terminology (such as mean, median, mode).

c. Demonstrates literal reading skills (such as recognizing main idea, details, sequencing, comparison and contrasting).

d. Demonstrates interpretative reading skills (such as predicting outcome, drawing conclusions, making generalizations).

e. Demonstrates critical reading skills (such as recognition of relevant and irrelevant information, propaganda techniques, and fallacies in reasoning).

f. Produces a logical summary interpretation of the results of research in professional material.

4. Demonstrate the ability to comprehend and work with fundamental mathematical concepts.

a. Adds, subtracts, multiplies, and divides whole numbers, decimals, and fractions.

b. Demonstrates the meaning and use of fractions and percents.

c. Represents and interprets data using charts, tables, graphs, and maps.

d. Solves measurement problems involving length, area, volume, capacity, weight, time and temperature, using U.S. customary and metric units.

e. Applies mathematical skills to solve real world problems.

f. Identifies geometric forms and relationships.

5. Demonstrate the ability to comprehend patterns of physical, social, and

academic development in students and to counsel students concerning their needs in these areas.

a. Demonstrates knowlege of basic principles of human growth and development.

b. Obtains knowledge of students through classroom tests, teacher observations, and student records to contribute to understanding of student needs.

c. Explains test data and other classroom evaluations to students in relation to their education needs.

d. Applies motivational techniques to encourage students to be achievement-oriented and goal-oriented.

e. Assists the student in relating achievements and interests to aptitude and ability.

f. Demonstrates knowledge of alternative school and community resources for students who have special needs.

g. Assists students in developing individual learning activities.

h. Develops student awareness of career opportunities using school and community resources.

6. Construct or assemble a classroom test to measure student performance according to criteria based on objectives.

a. Identifies uses of basic types of classroom tests and assessment techniques.

b. Identifies appropriate uses of norm-referenced and criterion-referenced testing.

c. Given an objective, specifies knowledge and skills to be assessed.

d. Selects appropriate assessment techniques to evaluate mastery of an

objective.

e. Determines limitations, constraints and requirements for administering tests.

f. Constructs and identifies test items and tasks that evaluate mastery of an objective.

g. Identifies criteria for standards of performance.

h. Assembles test components including test items, directions and scoring keys.

i. Evaluates and/or revises tests on the basis of validity, reliability and student responses.

7. Identify and/or develop a system for keeping records of class and individual student progress.

   a. Constructs a system for recording individual student knowledge and skill progress in a subject area.

   b. Identifies methods for reporting individual student progress in knowledge and skills in a subject area.

   c. Identifies methods for recording class progress in knowledge and skills in a subject area.

   d. Identifies methods for reporting class progress in knowledge and skills in a subject area.

   e. Demonstrates knowledge of the laws and policies governing the content and use of student records.

8. Demonstrate the ability to recognize and be aware of the instructional needs of exceptional students.

   a. Identifies the characteristics of exceptional students that have implications for modifying the learning environment.

b. Demonstrates awareness and appropriate use of educational programs, support services, personnel and other resources available to meet the needs of exceptional students.

c. Demonstrates the ability to identify and appropriately refer students who may be in need of exceptional student education.

d. Demonstrates awareness of the roles of the parent, teacher, and other professional personnel as members of the educational team responsible for planning, implementing, and evaluating the exceptional student's program.

e. Demonstrates the ability to recognize and/or use alternative instructional strategies to implement that portion of the exceptional student's program for which the teacher has responsibility.

f. Identifies and/or selects effective techniques and strategies for faciliating integration and social acceptance of exceptional students.

Competencies Most Effectively Evaluated by Think-Aloud or Interview Procedures or by Analysis of Products

9. Diagnose the entry level knowledge or skill to determine where instruction should begin.

a. Indicates what the student(s) knows and needs to know or be able to do to facilitate further learning.

b. Relates instructional elements (content, materials, activities) to diagnosed student needs.

c. Determines, by post test or other means, whether student learning needs have been satisfied.

10. Identify long-range goals for a given subject area.
    a. States intended student outcomes, both general and specific, that should result from instruction.
    b. Gives reason(s) for focusing on specific goals.
    c. Relates expected student outcomes to content, instructional activities, teaching-learning materials, instructional format or other instructional elements.
11. Construct and sequence related short-range objectives for a given subject area.
    a. Names one or more skills, concepts, facts, rules, principles, laws, or values to be taught during a period of instruction.
    b. Separates content to be taught into distinct elements or parts such as concepts and their exemplars, skills and their sequential steps.
    c. Judges appropriateness of content selected for a particular segment of instruction on the basis of specific criteria such as learner state, proper sequence, or time lines.
    d. Decides the order in which subject matter will be taught.
    e. Specifies or refers to the amount of subject matter to be taught during a segment of instruction.
12. Select, adapt, and/or develop instructional materials for a given set of instructional objectives and student learning needs.
    a. Names specific text pages or other types of materials to be used for instruction.
    b. Distinguishes characteristics of the material to be used for a given segment of instruction.
    c. Judges appropriateness of materials for a particular instructional

situation on the basis of specified criteria such as learner state, match with content or format of instruction, time required for completion, or availability.

d. Prepares materials or arranges to make instructional materials, to be used for a particular segment of instruction, readily available.

13. Select/develop and sequence related learning activities appropriate for a given set of instructional objectives and student learning needs.

a. States the activities in which she/he or the student will engage for a given period of instructional time.

b. Cites an order or pattern for a series of activities.

c. Breaks an activity into its component parts, specifying such things as sequential steps, how materials will be used, and teacher/student participation in the activity.

d. Specifies or refers to the amount of time to be spent on an instructional activity or the number of activities to be completed within a given instructional period.

e. Judges the appropriateness of an instructional activity on the basis of specified criteria such as learner state, match with content, instructional format, available time.

f. Shows attention to appropriate organization of the class for instructional activity.

Competencies Most Effectively Evaluated by Observation of Teacher Behavior in the Classroom

14. Demonstrate the ability to communicate information orally in a coherent, logical manner.
    a. Uses thematically connected discourse that leads to at least one point.
    b. Does not use discontinuous or garbled verbal behavior in which ideas are loosely associated.
    c. Does not use vague words to present information, or to answer questions, and for which the denotations are indeterminate (e.g., something, a little, you see, actually, things).
    d. Does not ask long and involved questions or multiple questions.
    e. Asks direct questions, one at a time, without rephrasing or giving additional information.
15. Establish rapport with students by using verbal and/or visual motivational devices.
    a. Uses marker expressions (e.g., note this, this is the most important point) to indicate that to which the student should give special attention.
    b. Uses marker techniques (e.g., underlining, colors, cartoons, etc.) in presenting information.
    c. Restates main points at spaced intervals.
    d. Makes zealous statements pointing out the task or activity is especially exciting or attractive (e.g., "This next exercise is going to be fun; I know you will enjoy it.").
    e. Makes statements indicating to students that an exercise or

activity will be challenging or hard to do.

f. Uses facial expressions, that give feedback about the teacher's positive affective state (pleasure, friendliness, interest, excitement, surprise).

g. Assumes posture and/or movement that indicates teacher's energy and enthusiasm.

h. Makes eye contacts without glaring, suspiciousness, or anger.

16. Present directions for carrying out an instructional activity.

a. Orients students to class work and engages them in academic activity.

b. Indicates movement of the lesson from one topic or activity to another by indicating where the lesson is and where it is going.

c. Explicates procedures for independent practice of lesson content.

d. Ascertains whether or not students understand what they are to do as seatwork/homework.

e. Alerts students as to when their seatwork/homework will be assessed.

17. Formulate standards for student behavior in the classroom.

a. States expectations about student conduct, giving rules or developing them with students, and illustrating rules for clarification.

b. Demonstrates rules and has students perform the behavior described by the rules.

c. Notes rule infractions and promptly tells student to desist and calls attention to the violated rule.

18. Identify causes of classroom misbehavior and employ a technique(s) for

correcting it.

a. Stops student deviant conduct before the deviant conduct spreads to other students or becomes more serious.

b. Selects the major disruption when two or more deviancies occur simultaneously.

c. Desists the student who caused the disruption, not a bystander.

d. Suggests different behavior to direct a student from deviant behavior.

e. Attends to a task and a disruption at the same time without affecting the on-going task.

f. Attends to two tasks at the same time without upsetting either one.

g. Gives clear desist by specifying who the deviant is, what he or she is doing wrong, and why this is improper conduct or what the proper conduct is.

h. Directs the deviant student to the task at hand as a desist is given.

i. Maintains movement smoothness and momentum of classroom activities.

j. Praises instances of appropriate conduct when the student has been misbehaving.

k. Explains to class that they encourage misconduct when they laugh at a misbehaving student.

l. Praises unobtrusively and quietly (almost unnoticeable by others).

m. Gives praise that reflects spontaneity, variety, simplicity, warmth, and meaningful content.

n. Does not make rough desists such as verbal expressions of anger or facial or bodily movements that express anger. (This avoids ripple effect.)

19. Demonstrate teaching skills which assist students in analyzing value claims.
    a. Specifies or leads students to specify the events, actions, conduct, or object to be evaluated and the evaluation term (e.g., good, bad, desirable, right).
    b. States and discusses criteria, or encourages students to state and discuss criteria by which to value an object, event, etc.
    c. Directs students in identifying and organizing facts that indicate whether the criteria are satisfied by the value object.
    d. Directs students in deciding whether they can accept the criteria as a generalized principle of valuation.
20. Establish a set of classroom routines and procedures for utilization and care of materials.
    a. Routinizes activities such as passing out papers, moving to get books, writing on the board, and has materials prepared, procedures worked out, and everything in order.
    b. Enforces rules and procedures to be followed by students who are tardy to class or who do not have their supplies, etc.
21. Identify and/or demonstrate behaviors which reflect a feeling for the dignity and worth of other people including those from other ethnic, cultural, linguistic, and economic groups.
    a. Accepts a student response by recognition of the comment.
    b. Modifies, rephrases, or applies a student response to redirect further discussion.
    c. Makes specific statements that give information about the value of a student's response or its implications.

22. Demonstrate instructional and social skills which assist students in developing a positive self-concept.
    a. Does not use rough desist; that is, utterances that express impatience or anger with the student, or facial or bodily behavior that expresses anger.
    b. Expresses desists that imply the teacher's warmth toward and feeling for students.
    c. Gives praise that reflects spontaneity, variety, warmth, and meaningful content.
    d. Utters expressions that set a warm and encouraging atmosphere for an examination.
    e. Comments positively on test results; that is, expresses approval of student responses to test item(s) or to discussion of responses.
23. Demonstrate instructional and social skills which assist students in interacting constructively with their peers.
    a. Asks different students to respond to questions or to react to another student's response.
    b. Comments positively on test results; that is, expresses approval of student responses to test item(s) or to discussion of responses.
    c. Modifies, rephrases, or applies a student's response to redirect further discussion.
    d. Asks questions that call for student's to explain, compare and contrast, evaluate, justify, or give evidence.
24. Demonstrate ability to stimulate and direct student thinking and to check student comprehension by questioning techniques.
    a. Asks question of individual students or the whole class about the lesson content to ascertain the level of understanding.

b. Asks the student to use information such as facts, definitions, names, and the like (student is not asked to give evidence to support his/her answer). These are low order questions.

c. Asks questions that require the student to explain, compare and contrast, evaluate, justify, and the like (student gives evidence or is askd to give evidence to support his/her answer). These are high order questions.

d. Does not ask non-academic questions for which any answer is acceptable (e.g., those that ask for expression of personal feelings or personal experiences).

e. Accepts only answers congruent with questions asked.

f. Asks a content question, then pauses before saying anything more and before soliciting an answer.

g. Asks question before calling on student in order to create suspense and group focus during recitation.

25. Demonstrate ability to provide practice to promote learning and retention.

a. Provides for choral practice where members of the class repeat examples of the task or psychomotor activities in unison.

26. Demonstrate ability to treat student talk in ways that encourage student participation and that maintain academic focus.

a. Accepts a student response by recognition of the student's comment.

b. Asks student to elaborate or justify his/her comment.

c. Modifies, rephrases, or applies a student response to redirect further discussion.

d. Does not respond to student questions or comments that disgress

from the learning task or could take the discussion to unrelated areas.

27. Demonstrate ability to use feedback procedures that give information to the student about the appropriateness of his/her response.
    a. Emits verbal or nonverbal behavior that signifies the student response is correct without elaboration.
    b. Makes specific statements that give information about the value of the response or its implications.
    c. Provides an explanation of a student error or gives a correction.
    d. Asks a different student to answer the question or to react to the response.
    e. Checks errors, or gives feedback, or grades seat/homework, or all of these.
28. Demonstrate ability to conduct review of subject matter.
    a. Rehearses previous lesson(s) (daily, weekly, monthly) at the beginning of a new lesson, or involves students in doing so.
    b. Recaps significant points of a preceding discussion as a concluding statement before moving to a new aspect of a topic or problem, or involves students in doing so.
    c. Makes a restatement or summary of the main points of discussion at the end of a lesson, or involves students in doing so.
29. Demonstrate ability to use class time efficiently.
    a. Begins classwork promptly.
    b. Manages transitions from one activity to another in a systematic, academically oriented way.
    c. Organizes the class to keep the lesson moving and provides

structure for those students who finish classwork early, thereby eliminating the necessity for students to wait for teacher approval.

30. Demonstrate ability to present forms of conceptual knowledge such as concepts, laws, and rules.

    a. Gives the verbal meaning of pivotal terms when beginning a new topic or some new aspect of a current topic, no analysis or examples being given.

    b. Names, describes, or depicts examples of a concept without giving either attributes or definition, or induces students to do so.

    c. Gives a definition and then names, describes or depicts examples that possess the attributes and examples that do not possess them, or induces students to do so.

    d. Gives a category and identifies the attributes for telling whether or not a given object or event belongs to the category, or induces students to do so.

    e. Emphasizes attributes that separate members of a family of concepts (pardon, probation, parole; or taxes, tariffs, fees), or encourages students to do so.

    f. Provides examples of a concept from which students infer the concept and its attributes.

    g. Identifies and analyzes the cause(s) of some phenomenon, or induces students to do so.

    h. Identifies and analyzes the effect of causal conditions, or directs students in an analysis of their effects.

    i. States causal principle or law, using linking words to connect

effect(s) to cause(s).

j. Directs students in using a causal principle to solve a problem or to explain a known effect.

k. Analyzes the kind of situation to which a given rule is applicable.

l. Gives a number of situations to help students learn to apply the rule.

31. Demonstrate ability to control the quality of vocal expression.

a. Does not talk in loud, noisy, or grating voice.

b. Does not talk in shrill, piercing, highly pitched voice.

c. Does not talk in monotone (fails to vary the intensity, rate, and volume of speech).

d. Does not talk too softly, almost inaudibly.

32. Demonstrate ability to use body language to express interest, excitement, joy and positive personal relations.

a. Uses facial expressions, that give feedback about his/her positive affective state (pleasure, friendliness, interest, excitement, surprise).

b. Does not display deadpan expression.

c. Makes facial expressions that give feedback about his/her negative affective state (displeasure, disapproval, anger, etc.).

d. Uses stance and movement to indicate energy and enthusiasm.

e. Uses eye contact, steadily and intensely without glaring, suspiciousness, or anger.

33. Demonstrate ability to give examinations in a manner to minimize anxiety and cheating and to provide appropriate feedback on test performance.

a. Informs students of the purpose of the test and content to be covered, how the test is relevant to them, and how the results will be used.

b. Instructs students in test format and how to take tests.

c. Arranges favorable test setting, creates an encouraging classroom atmosphere, and monitors the examination.

d. Uses test results to identify and correct errors and at the same time expresses approval of student responses to test items or discussion of answers.

04/02/84
CPT:A.AX1-17